

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): ACHIWA, Kyosuke
Serial No.: Not yet assigned
Filed: March 11, 2004
Title: CONTROL METHOD FOR STORAGE SYSTEM, STORAGE
SYSTEM, AND STORAGE DEVICE
Group: Not yet assigned

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

March 11, 2004

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on Japanese Patent Application No.(s) 2003-402994, filed December 2, 2003.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP



Carl I. Brundidge
Registration No. 29,621

CIB/alb
Attachment
(703) 312-6600

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 3 年 1 2 月 2 日
Date of Application:

出 願 番 号 特 願 2 0 0 3 - 4 0 2 9 9 4
Application Number:

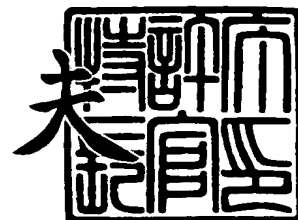
[ST. 10/C] : [J P 2 0 0 3 - 4 0 2 9 9 4]

出 願 人 株 式 会 社 日 立 製 作 所
Applicant(s):

2 0 0 4 年 1 月 3 0 日

特 許 庁 長 官
Commissioner,
Japan Patent Office

今 井 康 夫



出 証 番 号 出 証 特 2 0 0 4 - 3 0 0 4 6 0 3

【書類名】 特許願
【整理番号】 340301151
【提出日】 平成15年12月 2日
【あて先】 特許庁長官殿
【国際特許分類】 G06F 3/06
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 S A N
 ソリューション事業部内
 【氏名】 阿知和 恭介
【特許出願人】
 【識別番号】 000005108
 【氏名又は名称】 株式会社日立製作所
【代理人】
 【識別番号】 110000176
 【氏名又は名称】 一色国際特許業務法人
 【代表者】 一色 健輔
【手数料の表示】
 【予納台帳番号】 211868
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1

【書類名】 特許請求の範囲**【請求項 1】**

第一の情報処理装置と、

前記第一の情報処理装置と通信可能に接続され、前記第一の情報処理装置とクラスタを構成する第二の情報処理装置と、

前記第一の情報処理装置と通信可能に接続され、前記第一の情報処理装置から送信されてくるデータの入出力要求に応じて第一の記憶領域に対してデータの書き込み／読み込みを行う第一の記憶装置と、

前記第二の情報処理装置と通信可能に接続され、前記第二の情報処理装置から送信されてくるデータの入出力要求に応じて第二の記憶領域に対してデータの書き込み／読み込みを行う第二の記憶装置と、を含み、

前記第一の記憶装置と前記第二の記憶装置とは通信可能に接続され、

前記第一の記憶装置は、前記第一の記憶領域に書き込んだデータの複製を第二の記憶装置に送信し、前記データを受信した前記第二の記憶装置は、前記第二の記憶領域に前記データを書き込む処理である、第一の処理の実行中に、

前記第二の記憶装置は、前記第二の情報処理装置からフェールオーバーする旨の通知を受信すると、前記第一の記憶領域に書き込んだデータの複製が前記第二の記憶装置に未だ送信されておらず、前記データの複製が前記第二の記憶領域に書き込まれていないことを示す第一の情報を前記第一の記憶装置に要求し、

前記第二の記憶装置は、前記第一の記憶装置から前記第一の情報を受信すると、前記第二の情報処理装置にデータ入出力要求を受け付けることができる旨を通知し、

前記第二の記憶装置は、フェールオーバーした前記第二の情報処理装置から送信されてくるデータの読み出し要求を受信すると前記第一の情報を参照し、前記データの読み出し要求の対象となるデータが前記第一の記憶領域に記憶されていると判断した場合には、前記第一の記憶装置に前記データの読み出し要求の対象となるデータを要求し、当該要求に応じた前記第一の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを前記第二の情報処理装置に送信すること、

を特徴とするストレージシステムの制御方法。

【請求項 2】

請求項 1 に記載のストレージシステムの制御方法において、

前記第二の記憶装置は第三の記憶領域を備えており、

前記第二の記憶装置は、前記第一の処理の実行中に前記第二の情報処理装置からデータの書き込み要求を受信すると、前記データの書き込み要求の対象となるデータを前記第三の記憶領域に書き込み、

前記第一の処理が終了した場合には、前記第二の記憶装置は前記第三の記憶領域に書き込んだデータの複製を前記第二の記憶領域に書き込み、

前記第二の記憶装置は前記第三の記憶領域に書き込んだデータである書き込みデータを前記第一の記憶装置に送信し、前記書き込みデータを受信した前記第一の記憶装置は前記第一の記憶領域に前記書き込みデータを書き込む処理である、第二の処理を開始すること

を特徴とするストレージシステムの制御方法。

【請求項 3】

請求項 1 に記載のストレージシステムの制御方法において、

前記第二の記憶装置は第三の記憶領域を備えており、

前記第二の記憶装置は、前記第一の処理の実行中に前記第二の情報処理装置からデータの書き込み要求を受信すると、前記データの書き込み要求の対象となるデータを前記第三の記憶領域に書き込み、

前記第二の記憶装置は、前記第二の情報処理装置からデータの読み出し要求を受信すると、前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されているかを判断し、

前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されていないと判断した場合には、前記第二の記憶装置は前記第一の情報を参照して前記データの読み出し要求の対象となるデータが前記第二の記憶領域に記憶されているかを判断し、

前記データの読み出し要求の対象となるデータが前記第二の記憶領域に記憶されていないと判断した場合には、前記第二の記憶装置は前記データの読み出し要求の対象となるデータを前記第一の記憶装置に要求し、当該要求に応じた前記第一の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを受信すると、前記第二の記憶装置は前記データの読み出し要求の対象となるデータを前記第二の情報処理装置に送信すること、

を特徴とするストレージシステムの制御方法。

【請求項 4】

請求項 1 に記載のストレージシステムの制御方法において、

前記第二の記憶装置は第三の記憶領域を備えており、

前記第二の記憶装置は、前記第一の処理の実行中に前記第二の情報処理装置からデータの書き込み要求を受信すると、前記データの書き込み要求の対象となるデータを前記第三の記憶領域に書き込み、

前記第一の記憶装置は、前記第一の処理の実行中に前記第一の情報処理装置からフェールバックする旨の通知を受信すると、前記第三の記憶領域に書き込んだデータが前記第一の記憶装置に未だ送信されておらず、前記データが前記第一の記憶領域に書き込まれていないことを示す第二の情報を前記第二の記憶装置に要求し、

前記第一の記憶装置は、前記第二の記憶装置から前記第二の情報を受信すると、前記第一の情報処理装置にデータ入出力要求を受け付けることができる旨を通知し、

前記第二の記憶装置は、フェールバックした前記第一の情報処理装置から送信されてくるデータの読み出し要求を受信すると前記第二の情報を参照し、前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されていると判断した場合には、前記第二の記憶装置に前記データの読み出し要求の対象となるデータを要求し、当該要求に応じた前記第二の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを前記第一の情報処理装置に送信すること、

を特徴とするストレージシステムの制御方法。

【請求項 5】

請求項 4 に記載のストレージシステムの制御方法において、

前記第一の記憶装置は第四の記憶領域を備えており、

前記第一の記憶装置は、前記第一の処理の実行中に前記第一の情報処理装置からデータの書き込み要求を受信すると、前記データの書き込み要求の対象となるデータを前記第四の記憶領域に書き込み、

前記第一の処理が終了した場合には、前記第二の記憶装置は前記第三の記憶領域に書き込んだデータの複製を前記第二の記憶領域に書き込み、

前記第二の記憶装置は、前記第三の記憶領域に書き込んだデータである書き込みデータを前記第一の記憶装置に送信し、前記書き込みデータを受信した前記第一の記憶装置は前記第一の記憶領域に前記書き込みデータを書き込む処理である、第二の処理を開始し、

前記第二の処理が終了した場合には、前記第一の記憶装置は前記第四の記憶領域に書き込んだデータの複製を前記第一の記憶領域に書き込み、

前記第一の記憶装置は、前記第四の記憶領域に書き込んだデータである書き込みデータを前記第二の記憶装置に送信し、前記書き込みデータを受信した前記第二の記憶装置は前記第二の記憶領域に前記書き込みデータを書き込む処理である、第三の処理を開始すること、

を特徴とするストレージシステムの制御方法。

【請求項 6】

請求項 4 に記載のストレージシステムの制御方法において、

前記第一の記憶装置は第四の記憶領域を備えており、

前記第一の記憶装置は、前記第一の処理の実行中に前記第一の情報処理装置からデータの書き込み要求を受信すると、前記データの書き込み要求の対象となるデータを前記第四の記憶領域に書き込み、

前記第一の記憶装置は、前記第一の情報処理装置からデータの読み出し要求を受信すると、前記データの読み出し要求の対象となるデータが前記第四の記憶領域に記憶されているかを判断し、

前記データの読み出し要求の対象となるデータが前記第四の記憶領域に記憶されていないと判断した場合には、前記第一の記憶装置は、前記第二の情報を参照して前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されているかを判断し、

前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されていると判断した場合には、前記第一の記憶装置は前記データを前記第二の記憶装置に要求し、当該要求に応じた前記第二の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを受信すると、前記第一の記憶装置は前記データの読み出し要求の対象となるデータを前記第一の情報処理装置に送信すること、

を特徴とするストレージシステムの制御方法。

【請求項 7】

請求項 1 に記載のストレージシステムの制御方法において、

前記第一の情報処理装置は、前記第一の記憶装置に前記データの書き込み要求を送信する時刻を監視しており、

前記第一の情報処理装置は、前記データの書き込み要求を前記第一の記憶装置に送信するとともに、当該送信時刻の情報を送信し、

前記第一の記憶装置は、前記第一の情報処理装置から順次送信されてくる前記データの書き込み要求の対象となるデータをそれぞれ送信時刻の情報と対応づけて管理し、

前記第一の記憶装置は、前記第一の記憶領域に記憶した前記データを古いものから順に前記第二の記憶装置に送信すること、

を特徴とするストレージシステムの制御方法。

【請求項 8】

第一の情報処理装置と、

前記第一の情報処理装置と通信可能に接続され、前記第一の情報処理装置とクラスタを構成する第二の情報処理装置と、

前記第一の情報処理装置と通信可能に接続され、前記第一の情報処理装置から送信されてくるデータの入出力要求に応じて第一の記憶領域に対してデータの書き込み／読み込みを行う第一の記憶装置と、

前記第二の情報処理装置と通信可能に接続され、前記第二の情報処理装置から送信されてくるデータの入出力要求に応じて第二の記憶領域に対してデータの書き込み／読み込みを行う第二の記憶装置と、を含み、

前記第一の記憶装置と前記第二の記憶装置とは通信可能に接続され、

前記第一の記憶装置が前記第一の記憶領域に書き込んだデータの複製を前記第二の記憶装置に送信し、前記データを受信した前記第二の記憶装置が前記第二の記憶領域に前記データを書き込む処理である、第一の処理を実行する複製処理部を備え、

前記第二の記憶装置は、チャネル制御部を備え、

前記チャネル制御部は、前記第一の処理の実行中に前記第二の情報処理装置からフェールオーバーする旨の通知を受信すると、前記第一の記憶領域に書き込んだデータの複製が前記第二の記憶装置に未だ送信されておらず、前記データの複製が前記第二の記憶領域に描き込まれていないことを示す第一の情報を前記第一の記憶装置に要求し、

前記チャネル制御部は、前記第一の記憶装置から前記第一の情報を受信すると、前記第二の情報処理装置にデータ入出力要求を受け付けることができる旨を通知し、

前記チャネル制御部は、フェールオーバーした前記第二の情報処理装置から送信されてくるデータの読み出し要求を受信すると前記第一の情報を参照し、前記データの読み出し要求の対象となるデータが前記第一の記憶領域に記憶されていると判断した場合には、前記

第一の記憶装置に前記データの読み出し要求の対象となるデータを要求し、当該要求に応じた前記第一の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを前記第二の情報処理装置に送信すること、
を特徴とするストレージシステム。

【請求項 9】

請求項 8 に記載のストレージシステムにおいて、
前記第二の記憶装置は、第三の記憶領域とディスク制御部とを備え、
前記チャネル制御部は、前記第一の処理の実行中に前記第二の情報処理装置からデータの書き込み要求を受信し、
前記ディスク制御部は、前記データの書き込み要求の対象となるデータを前記第三の記憶領域に書き込み、
前記ディスク制御部は、前記第一の処理が終了した場合には、前記第三の記憶領域に書き込んだデータの複製を前記第二の記憶領域に書き込み、
前記複製処理部は、前記第二の記憶装置が前記第三の記憶領域に書き込んだデータを前記第一の記憶装置に送信し、前記データを受信した前記第一の記憶装置が前記第一の記憶領域に前記データを書き込む処理である、第二の処理を開始すること、
を特徴とするストレージシステム。

【請求項 10】

請求項 8 に記載のストレージシステムにおいて、
前記第二の記憶装置は、第三の記憶領域とディスク制御部とを備え、
前記チャネル制御部は、前記第一の処理の実行中に前記第二の情報処理装置からデータの書き込み要求を受信し、
前記ディスク制御部は、前記データの書き込み要求の対象となるデータを前記第三の記憶領域に書き込み、
前記チャネル制御部は、前記第二の情報処理装置からデータの読み出し要求を受信し、
前記チャネル制御部は、前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されているかを判断し、
前記チャネル制御部は、前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されていないと判断した場合には、前記第一の情報を参照して前記データの読み出し要求の対象となるデータが前記第二の記憶領域に記憶されているかを判断し、
前記チャネル制御部は、前記データの読み出し要求の対象となるデータが前記第二の記憶領域に記憶されていないと判断した場合には、前記データの読み出し要求の対象となるデータを前記第一の記憶装置に要求し、当該要求に応じた前記第一の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを受信すると、前記データの読み出し要求の対象となるデータを前記第二の情報処理装置に送信すること、
を特徴とするストレージシステム。

【請求項 11】

請求項 8 に記載のストレージシステムにおいて、
前記第二の記憶装置は、第三の記憶領域とディスク制御部とを備え、
前記第一の記憶装置は、チャネル制御部を備え、
前記第二の記憶装置のチャネル制御部は、前記第一の処理の実行中に前記第二の情報処理装置からデータの書き込み要求を受信し、
前記ディスク制御部は、前記データの書き込み要求の対象となるデータを前記第三の記憶領域に書き込み、
前記第一の記憶装置のチャネル制御部は、記第一の処理の実行中に前記第一の情報処理装置からフェールバックする旨の通知を受信すると、前記第三の記憶領域に書き込んだデータが前記第一の記憶装置に未だ送信されておらず、前記データが前記第一の記憶領域に書き込まれていないことを示す第二の情報を前記第二の記憶装置に要求し、
前記第一の記憶装置のチャネル制御部は、前記第二の記憶装置から前記第二の情報を受信すると、前記第一の情報処理装置にデータ入出力要求を受け付けることができる旨を通

知し、

前記第一の記憶装置のチャネル制御部は、フェールバックした前記第一の情報処理装置から送信されてくるデータの読み出し要求を受信すると、前記第二の情報を参照して前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されていると判断した場合には、前記第二の記憶装置に前記データの読み出し要求の対象となるデータを要求し、当該要求に応じた前記第二の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを前記第一の情報処理装置に送信すること、

を特徴とするストレージシステム。

【請求項 12】

請求項 11 に記載のストレージシステムにおいて、

前記第一の記憶装置は、第四の記憶領域とディスク制御部とを備え、

前記第一の記憶装置のチャネル制御部は、前記第一の処理の実行中に前記第一の情報処理装置からデータの書き込み要求を受信し、

前記第一の記憶装置のディスク制御部は、前記データの書き込み要求の対象となるデータを前記第四の記憶領域に書き込み、

前記第二の記憶装置のディスク制御部は、前記第一の処理が終了した場合に、前記第三の記憶領域に書き込んだデータの複製を前記第二の記憶領域に書き込み、

前記複製処理部は、前記第二の記憶装置が前記第三の記憶領域に書き込んだデータを前記第一の記憶装置に送信し、前記データを受信した前記第一の記憶装置が前記第一の記憶領域に前記データを書き込む処理である、第二の処理を開始し、

前記第一の記憶装置のディスク制御部は、前記第二の記憶装置から受信した前記第三の記憶領域に書き込まれているデータを前記第一の記憶領域に書き込み、

前記第一の記憶装置のディスク制御部は、前記第二の処理が終了した場合に、前記第四の記憶領域に書き込んだデータの複製を前記第一の記憶領域に書き込み、

前記複製処理部は、前記第一の記憶装置が前記第四の記憶領域に書き込んだデータを前記第二の記憶装置に送信し、前記データを受信した前記第二の記憶装置が前記第二の記憶領域に前記データを書き込む処理である、第三の処理を開始すること、

を特徴とするストレージシステム。

【請求項 13】

請求項 11 に記載のストレージシステムにおいて、

前記第一の記憶装置は、第四の記憶領域とディスク制御部とを備え、

前記第一の記憶装置のチャネル制御部は、前記第一の処理の実行中に前記第一の情報処理装置からデータの書き込み要求を受信し、

前記第一の記憶装置のディスク制御部は、前記データの書き込み要求の対象となるデータを前記第四の記憶領域に書き込み、

前記第一の記憶装置のチャネル制御部は、前記第一の情報処理装置からデータの読み出し要求を受信し、

前記第一の記憶装置のチャネル制御部は、前記データの読み出し要求の対象となるデータが前記第四の記憶領域に記憶されているかを判断し、

前記第一の記憶装置のチャネル制御部は、前記データの読み出し要求の対象となるデータが前記第四の記憶領域に記憶されていないと判断した場合には、前記第二の情報を参照して前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されているかを判断し、

前記第一の記憶装置のチャネル制御部は、前記データの読み出し要求の対象となるデータが前記第二の記憶領域に記憶されていると判断した場合には、前記データの読み出し要求の対象となるデータを前記第二の記憶装置に要求し、当該要求に応じた前記第二の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを受信し、前記データの読み出し要求の対象となるデータを前記第一の情報処理装置に送信すること、

を特徴とするストレージシステム。

【請求項 14】

請求項 8 に記載のストレージシステムにおいて、

前記第一の情報処理装置は、

前記第一の記憶装置に前記データの書き込み要求を送信する時刻を監視するタイマーと

、
前記データの書き込み要求を前記第一の記憶装置に送信するとともに、当該送信時刻の情報を送信するプロセッサと、を備え、

前記第一の記憶装置は、

前記第一の情報処理装置から順次送信されてくる前記データの書き込み要求の対象となるデータをそれぞれ送信時刻の情報と対応づけて管理する前記チャネル制御部と、

前記第一の記憶領域に記憶されているデータを古いものから順に前記第二の記憶装置に送信する前記チャネル制御部と、を備える

ことを特徴とするストレージシステム。

【請求項 15】

第一の情報処理装置と、

前記第一の情報処理装置と通信可能に接続され、前記第一の情報処理装置とクラスタを構成する第二の情報処理装置と、

前記第一の情報処理装置と通信可能に接続され、前記第一の情報処理装置から送信されてくるデータの入出力要求に応じて第一の記憶領域に対してデータの書き込み／読み込みを行う他の記憶装置と、

前記第二の情報処理装置と通信可能に接続され、前記第二の情報処理装置から送信されてくるデータの入出力要求に応じて第二の記憶領域に対してデータの書き込み／読み込みを行う記憶装置と、を含み、

前記他の記憶装置と前記記憶装置とは通信可能に接続され、

前記他の記憶装置が前記第一の記憶領域に書き込んだデータの複製を前記記憶装置に送信し、前記データを受信した前記記憶装置が前記第二の記憶領域に前記データを書き込む処理である、第一の処理を実行する複製処理部を備えるストレージシステムの記憶装置であって、

前記記憶装置は、チャネル制御部を備え、

前記チャネル制御部は、前記第一の処理の実行中に前記第二の情報処理装置からフェールオーバーする旨の通知を受信すると、前記第一の記憶領域に書き込んだデータの複製が前記記憶装置に未だ送信されておらず、前記データの複製が前記第二の記憶領域に描き込まれていないことを示す第一の情報を前記他の記憶装置に要求し、

前記チャネル制御部は、前記他の記憶装置から前記第一の情報を受信すると、前記第二の情報処理装置にデータ入出力要求を受け付けることができる旨を通知し、

前記チャネル制御部は、フェールオーバーした前記第二の情報処理装置から送信されてくるデータの読み出し要求を受信すると前記第一の情報を参照し、前記データの読み出し要求の対象となるデータが前記第一の記憶領域に記憶されていると判断した場合には、前記他の記憶装置に前記データの読み出し要求の対象となるデータを要求し、当該要求に応じた前記他の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを前記第二の情報処理装置に送信すること、

を特徴とする記憶装置。

【請求項 16】

請求項 15 に記載の記憶装置において、

第三の記憶領域とディスク制御部とを備え、

前記チャネル制御部は、前記第一の処理の実行中に前記第二の情報処理装置からデータの書き込み要求を受信し、

前記ディスク制御部は、前記データの書き込み要求の対象となるデータを前記第三の記憶領域に書き込み、

前記ディスク制御部は、前記第一の処理が終了した場合には、前記第三の記憶領域に書き込んだデータの複製を前記第二の記憶領域に書き込み、

前記チャネル制御部は、前記第三の記憶領域に書き込んだデータを前記他の記憶装置に送信すること、

を特徴とする記憶装置。

【請求項 17】

請求項 15 に記載の記憶装置において、

第三の記憶領域とディスク制御部とを備え、

前記チャネル制御部は、前記第一の処理の実行中に前記第二の情報処理装置からデータの書き込み要求を受信し、

前記ディスク制御部は、前記データの書き込み要求の対象となるデータを前記第三の記憶領域に書き込み、

前記チャネル制御部は、前記第二の情報処理装置からデータの読み出し要求を受信し、

前記チャネル制御部は、前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されているかを判断し、

前記チャネル制御部は、前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されていないと判断した場合には、前記第一の情報を参照して前記データの読み出し要求の対象となるデータが前記第二の記憶領域に記憶されているかを判断し、

前記チャネル制御部は、前記データの読み出し要求の対象となるデータが前記第二の記憶領域に記憶されていないと判断した場合には、前記データの読み出し要求の対象となるデータを前記他の記憶装置に要求し、当該要求に応じた前記第一の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを受信すると、前記データの読み出し要求の対象となるデータを前記第二の情報処理装置に送信すること、

を特徴とするストレージシステム。

【請求項 18】

第一の情報処理装置と、

前記第一の情報処理装置と通信可能に接続され、前記第一の情報処理装置とクラスタを構成する第二の情報処理装置と、

前記第一の情報処理装置と通信可能に接続され、前記第一の情報処理装置から送信されてくるデータの入出力要求に応じて第一の記憶領域に対してデータの書き込み／読み込みを行う記憶装置と、

前記第二の情報処理装置と通信可能に接続され、前記第二の情報処理装置から送信されてくるデータの入出力要求に応じて第二の記憶領域に対してデータの書き込み／読み込みを行う他の記憶装置と、を含み、

前記記憶装置と前記他の記憶装置とは通信可能に接続され、

前記記憶装置が前記第一の記憶領域に書き込んだデータの複製を前記他の記憶装置に送信し、前記データを受信した前記他の記憶装置が前記第二の記憶領域に前記データを書き込む処理である、第一の処理を実行する複製処理部を備え、

前記他の記憶装置は、第三の記憶領域を備え、

前記他の記憶装置は、前記第一の処理の実行中に前記第二の情報処理装置からデータの書き込み要求を受信し、前記データの書き込み要求の対象となるデータを前記第三の記憶領域に書き込む、ストレージシステムの記憶装置であって、

前記記憶装置は、チャネル制御部を備え、

前記チャネル制御部は、記第一の処理の実行中に前記第一の情報処理装置からフェールバックする旨の通知を受信すると、前記第三の記憶領域に書き込んだデータが前記第一の記憶装置に未だ送信されておらず、前記データが前記第一の記憶領域に書き込まれていないことを示す第二の情報を前記第二の記憶装置に要求し、

前記チャネル制御部は、前記第二の記憶装置から前記第二の情報を受信すると、前記第一の情報処理装置にデータ入出力要求を受け付けることができる旨を通知し、

前記チャネル制御部は、フェールバックした前記第一の情報処理装置から送信されてくるデータの読み出し要求を受信すると、前記第二の情報を参照して前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されていると判断した場合には、前

記第二の記憶装置に前記データの読み出し要求の対象となるデータを要求し、当該要求に応じた前記第二の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを前記第一の情報処理装置に送信すること、
を特徴とする記憶装置。

【請求項 19】

請求項 18 に記載のストレージシステムにおいて、
前記記憶装置は、第四の記憶領域とディスク制御部とを備え、
前記チャネル制御部は、前記第一の処理の実行中に前記第一の情報処理装置からデータの書き込み要求を受信し、
前記ディスク制御部は、前記データの書き込み要求の対象となるデータを前記第四の記憶領域に書き込み、
前記チャネル制御部は、前記他の記憶装置から送信されてくる前記第三の記憶領域に書き込まれているデータを受信し、
前記ディスク制御部は、前記第三の記憶領域に書き込まれているデータを前記第一の記憶領域に書き込み、
前記ディスク制御部は、前記第三の記憶領域に書き込まれている全てのデータを書き込むと、前記第四の記憶領域に書き込んだデータの複製を前記第一の記憶領域に書き込み、
前記チャネル制御部は、前記第四の記憶領域に書き込んだデータを前記第二の記憶装置に送信すること、
を特徴とする記憶装置。

【請求項 20】

請求項 18 に記載のストレージシステムにおいて、
前記記憶装置は、第四の記憶領域とディスク制御部とを備え、
前記チャネル制御部は、前記第一の処理の実行中に前記第一の情報処理装置からデータの書き込み要求を受信し、
前記ディスク制御部は、前記データの書き込み要求の対象となるデータを前記第四の記憶領域に書き込み、
前記チャネル制御部は、前記第一の情報処理装置からデータの読み出し要求を受信し、
前記チャネル制御部は、前記データの読み出し要求の対象となるデータが前記第四の記憶領域に記憶されているかを判断し、
前記チャネル制御部は、前記データの読み出し要求の対象となるデータが前記第四の記憶領域に記憶されていないと判断した場合には、前記第二の情報を参照して前記データの読み出し要求の対象となるデータが前記第三の記憶領域に記憶されているかを判断し、
前記チャネル制御部は、前記データの読み出し要求の対象となるデータが前記第二の記憶領域に記憶されていると判断した場合には、前記データの読み出し要求の対象となるデータを前記他の記憶装置に要求し、当該要求に応じた前記他の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを受信し、前記データの読み出し要求の対象となるデータを前記第一の情報処理装置に送信すること、
を特徴とする記憶装置。

【書類名】明細書**【発明の名称】** ストレージシステムの制御方法、ストレージシステム、及び記憶装置**【技術分野】****【0001】**

本発明は、ストレージシステムの制御方法、ストレージシステム、及び記憶装置に関する。

【背景技術】**【0002】**

ストレージシステムにおける災害復旧（ディザスタリカバリ）が注目されている。ディザスタリカバリを実現する技術として、複製元の記憶領域のデータの複製を複製先の記憶領域においても管理する技術（リモートコピー）が知られている（例えば、特許文献1～3参照）。この技術により、複製元の記憶装置にアクセスする情報処理装置の障害時に、複製先の記憶装置のデータを使用することで、前記情報処理装置で行われていた処理を、複製先の記憶装置にアクセスする他の情報処理装置に引き継がせることができる。

【特許文献1】 特開2001-337939号公報**【特許文献2】** 米国特許出願公開第2003/51111号明細書**【特許文献3】** 米国特許第6591351号明細書**【発明の開示】****【発明が解決しようとする課題】****【0003】**

ところで、複製元の記憶装置にアクセスする情報処理装置から複製先の記憶装置にアクセスする情報処理装置への上記引き継ぎに際し、複製元の記憶装置から複製先の記憶装置へのデータのセットアップが完了していないことがある。この場合、上記セットアップが完了するまで複製先の記憶装置にアクセスする情報処理装置を待機させることとなり、場合によっては情報処理装置側でタイムアウトを生じるなど、スムーズな引き継ぎが行えないことがある。

【0004】

本発明は、情報処理装置間での処理の引き継ぎをスムーズに行うことができる、ストレージシステムの制御方法、ストレージシステム、及び記憶装置を提供することを目的とする。

【課題を解決するための手段】**【0005】**

上記課題を解決するために、本発明のストレージシステムの制御方法は、第一の情報処理装置と、前記第一の情報処理装置と通信可能に接続され、前記第一の情報処理装置とクラスタを構成する第二の情報処理装置と、前記第一の情報処理装置と通信可能に接続され、前記第一の情報処理装置から送信されてくるデータの入出力要求に応じて第一の記憶領域に対してデータの書き込み／読み込みを行う第一の記憶装置と、前記第二の情報処理装置と通信可能に接続され、前記第二の情報処理装置から送信されてくるデータの入出力要求に応じて第二の記憶領域に対してデータの書き込み／読み込みを行う第二の記憶装置と、を含み、前記第一の記憶装置と前記第二の記憶装置とは通信可能に接続され、前記第一の記憶装置は、前記第一の記憶領域に書き込んだデータの複製を第二の記憶装置に送信し、前記データを受信した前記第二の記憶装置は、前記第二の記憶領域に前記データを書き込む処理である、第一の処理の実行中に、前記第二の記憶装置は、前記第二の情報処理装置からフェールオーバーする旨の通知を受信すると、前記第一の記憶領域に書き込んだデータの複製が前記第二の記憶装置に未だ送信されておらず、前記データの複製が前記第二の記憶領域に書き込まれていないことを示す第一の情報を前記第一の記憶装置に要求し、前記第二の記憶装置は、前記第一の記憶装置から前記第一の情報を受信すると、前記第二の情報処理装置にデータ入出力要求を受け付けることができる旨を通知し、前記第二の記憶装置は、フェールオーバーした前記第二の情報処理装置から送信されてくるデータの読み出し要求を受信すると前記第一の情報を参照し、前記データの読み出し要求の対象となるデ

ータが前記第一の記憶領域に記憶されていると判断した場合には、前記第一の記憶装置に前記データの読み出し要求の対象となるデータを要求し、当該要求に応じた前記第一の記憶装置から送信されてくる前記データの読み出し要求の対象となるデータを前記第二の情報処理装置に送信することとする。

【0006】

その他、本願が開示する課題、及びその解決方法は、発明の実施の形態の欄、及び図面により明らかにされる。

【発明の効果】

【0007】

本発明によれば、情報処理装置間の処理の引き継ぎをスムーズに行うことができる、ストレージシステムの制御方法、ストレージシステム、及び、記憶装置を提供することができる。

【発明を実施するための最良の形態】

【0008】

図1は本発明の一実施例として説明するストレージシステム90の全体構成のブロック図を示す。

【0009】

ストレージシステム90は、第一の情報処理装置10と、これと通信可能に接続されている第二の情報処理装置20と、第一の情報処理装置10と通信可能に接続されている第一の記憶装置30と、第二の情報処理装置20と通信可能に接続されている第二の記憶装置40とを含んで構成される。ストレージシステム90は、例えば、銀行のオンラインや経理等の業務、商社、物流会社などにおける在庫管理、鉄道会社や航空会社における座席予約などに運用されるシステムである。このストレージシステム90は、地震・火災・台風・落雷・テロなどに対するディザスタリカバリの実現のために構築される。

【0010】

第一の記憶装置30と第二の記憶装置40は第一のネットワーク80を介して通信可能に接続されている。第一のネットワーク80は、例えば、ギガビットイーサネット（登録商標）、ATM（AsynchroN Ous Transfer Mode）、公衆回線などである。

【0011】

情報処理装置10、20と記憶装置30、40とは第二のネットワーク50、60を介して通信可能に接続されている。第二のネットワーク50、60は、例えば、SAN（Storage Area Network）である。SANは、記憶装置30、40のディスクドライブが提供する記憶領域におけるデータの管理単位であるブロックを単位として情報処理装置10、20との間でデータの授受を行うためのネットワークである。なお、第二のネットワーク50、60は、LAN（Local Area Network）や、iSCSI（Internet Small Computer System Interface）、Fibre Channel、ESCON（Enterprise Systems Connection）（登録商標）、FICON（Fibre Connection）（登録商標）などであってもよい。

【0012】

各情報処理装置10、20は第三のネットワーク70を介して通信可能に接続されている。第三のネットワーク70は、例えば、LAN（Local Area Network）などである。

【0013】

図2は記憶装置30、40の一例として説明するディスクアレイ装置の具体的な構成を示す。なお、記憶装置30、40は、ディスクアレイ装置以外にも、例えば、半導体記憶装置などであってもよい。

【0014】

ディスクアレイ装置は、記憶制御装置100とディスクドライブ108などを備えている。記憶制御装置100は、チャネル制御部101、リモート通信インタフェース102、ディスク制御部103、共有メモリ104、キャッシュメモリ105、これらの間を通信可能に接続するクロスバススイッチなどで構成されるスイッチング制御部106などを備

えて構成される。

キャッシュメモリ 105 は、主としてチャネル制御部 101 とディスク制御部 103 との間で授受されるデータを一時的に記憶するために用いられる。

【0015】

ディスク制御部 103 は、チャネル制御部 101 により共有メモリ 104 に書き込まれたデータの入出力要求を読み出してそのデータの入出力要求に指定されているコマンド（例えば、SCSI (Small Computer System Interface) 規格のコマンド）に従ってディスクドライブ 108 にデータの書き込みや読み出しなどの処理を実行する。ディスク制御部 103 は、ディスクドライブ 108 をいわゆる RAID (Redundant Array of Inexpensive Disks) 方式に規定される RAID レベル（例えば、0, 1, 5）で制御する機能を備えることもある。

【0016】

ディスクドライブ 108 は、例えば、ハードディスク装置である。ディスクドライブ 108 はディスクアレイ装置と一体型とすることもできるし、別体とすることもできる。ディスクドライブ 108 により提供される記憶領域は、物理ボリューム又はその物理ボリューム上に論理的に設定される論理ボリュームを単位として管理されている。ディスクドライブ 108 へのデータの書き込みや読み出しは、論理ボリュームに付与される識別子を指定して行なうことができる。

【0017】

管理端末 107 はディスクアレイ装置やディスクドライブ 108 を保守・管理するためのコンピュータである。例えば、チャネル制御部 101 やディスク制御部 103 において実行されるソフトウェアやパラメータの変更や、ディスクドライブ 108 の構成の設定や、論理ボリュームの管理又は設定（容量管理や容量拡張・縮小、情報処理装置 10, 20 の割り当て等）等は、管理端末 107 からの指示により行われる。管理端末 107 はディスクアレイ装置に内蔵される形態とすることもできるし、別体とすることもできる。

【0018】

リモート通信インタフェース 102 は、他の記憶装置 30, 40 とデータ伝送をするための通信インタフェース（チャネルエクステンダ）であり、後述するリモートコピーにおける複製データの伝送はこのリモート通信インタフェース 102 を介して行われる。リモート通信インタフェース 102 は、チャネル制御部 101 のインタフェース（例えば、Fibre Channel、ESCON（登録商標）、FICON（登録商標）などのインタフェース）を第一のネットワーク 80 の通信方式に変換する。これにより他の記憶装置 30, 40 との間でのデータ伝送が実現される。

【0019】

なお、ディスクアレイ装置は、以上に説明した構成のもの以外にも、例えば、NFS (Network File System) などのプロトコルにより情報処理装置 10, 20 からファイル名指定によるデータ入出力要求を受け付けるように構成された NAS (Network Attached Storage) として機能するものなどであってもよい。

【0020】

共有メモリ 104 はチャネル制御部 101 とディスク制御部 103 の両方からアクセスが可能である。データ入出力要求コマンドの受け渡しに利用される他、記憶装置 30, 40 やディスクドライブ 108 の管理情報等が記憶される。

【0021】

図 3 は本実施の形態に係る情報処理装置 10, 20 の構成の一例を示すブロック図である。

情報処理装置 10, 20 は、CPU（プロセッサ）110、メモリ 120、ポート 130、記憶装置 140、記録媒体読取装置 160、入力装置 170、出力装置 180、タイマー 200などを備える。

【0022】

CPU 110 は情報処理装置 10, 20 の全体の制御を司るものであり、メモリ 120

に格納されたプログラムを実行することにより各種機能を実現し、後述する情報処理装置 10, 20 の処理を行う。例えば上述した銀行の自動預金預け払いシステムを実現するための機能や航空機の座席予約システムを実現するための機能等である。

記録媒体読取装置 160 は、記録媒体 190 に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ 120 や記憶装置 140 に格納される。記録媒体 190 としてはフレキシブルディスクや CD-ROM、半導体メモリ等を用いることができる。記録媒体読取装置 160 は情報処理装置 10, 20 に内蔵されている形態とすることもできるし、外付されている形態とすることもできる。

入力装置 170 はオペレータ等による情報処理装置 10, 20 へのデータ入力等のために用いられる。入力装置 170 は、例えばキーボードやマウス等である。出力装置 180 は情報を外部に出力するための装置である。出力装置 180 は、例えばディスプレイやプリンタ等である。ポート 130 は記憶装置 30, 40 と通信を行うための装置である。また他の情報処理装置 10, 20 との間で通信を行うために使用することもできる。従って、情報処理装置 10, 20 は、メモリ 120 や記憶装置 140 に格納されているプログラムやデータを、ポート 130 を介して他の情報処理装置 10, 20 から受信して、メモリ 120 に格納するようにすることもできる。

タイマー 200 は、本実施の形態におけるいくつかの処理を行う時間を監視している。タイマー 200 は、例えば、ハードウェアのタイマーや、ソフトウェアのタイマーなどである。このタイマー 200 により、情報処理装置 10, 20 はあらかじめ設定された時間内にデータの処理やデータの伝送が終わらないときにタイムアウト異常が発生したと判断して、処理や通信を打ち切って回復処理を実行することが可能となる。従って、タイムアウトを放置することにより起こるシステムダウンを防止することができるようになる。バス 150 はこれらの構成を相互に接続する。

【0023】

図 4 は本実施の形態に係るチャネル制御部 101 の構成の一例を示すブロック図である。

チャネル制御部 101 は、CPU 211、キャッシュメモリ 212、制御メモリ 213、ポート 215、バス 216 などを備えている。

【0024】

CPU 211 は、チャネル制御部 101 の全体の制御を司るものであり、制御メモリ 213 に格納されたプログラムを実行することにより、後述するチャネル制御部 101 の各処理を行う。制御メモリ 213 に格納された制御プログラム（複製処理部）214 は、各記憶装置 30, 40 の CPU 211 によって実行されることにより、後述するリモートコピーを実行する。キャッシュメモリ 212 は情報処理装置 10, 20 との間で授受されるデータやコマンド等を一時的に格納するためのメモリである。ポート 215 は情報処理装置 10, 20 との間の通信や他の記憶装置 30, 40 との間の通信を行うための通信インタフェースである。バス 216 はこれらの構成を相互に接続する。

【0025】

図 5 は本発明の一実施例として説明するストレージシステムの概略構成を示す図である。

【0026】

第一の情報処理装置 10 と第二の情報処理装置 20 には、高可用性 (High Availability) を図ることを目的としてクラスタ 310 を実現するためのハードウェアやソフトウェアなどが導入されている。本実施の形態においてクラスタとは、主にフェールオーバー型クラスタを意味する。フェールオーバー型クラスタを構成するとは、2 台あるいはそれ以上の情報処理装置を同時に動作させ、1 台を主（プライマリ）として、他の情報処理装置を副（セカンダリ）として利用し、何らかの原因で主の情報処理装置において障害が発生した時に、他の情報処理装置が主の情報処理装置で行っていた処理を引き継ぐように構成することをいう。この情報処理装置 10, 20 のクラスタ 310 により、情報処理装置 10, 20 は、他の情報処理装置 10, 20 の障害を第三のネットワーク 70 を介して互いに監

視することが可能になっている。また、情報処理装置 10, 20 のクラスタ 310 により、情報処理装置 10, 20 は、他の情報処理装置 10, 20 の障害を検知すると他の情報処理装置 10, 20 で行っていた処理 300 を引き継ぐ（フェールオーバーする）ことが可能になっている。また、情報処理装置 10, 20 のクラスタ 310 により、第一の情報処理装置 10 の障害が復旧したと判断した場合には、第一の情報処理装置 10 は、第二の情報処理装置 20 で行っていた処理 300 を引き継ぐ（フェールバックする）ことが可能になっている。なお、第一の情報処理装置 10 が第二の情報処理装置 20 の障害を検知した場合や、管理端末 107 からフェールバック要求があった場合などにフェールバックするように設定してもよい。

【0027】

情報処理装置 10, 20 は、エージェント 320 により記憶装置 30, 40 に対してデータの読み出し要求やデータの書き込み要求を送信することができ、記憶装置 30, 40 から読み出しデータやデータの書き込み完了通知を受信することができる。

【0028】

第一の記憶装置 30 は、差分管理テーブル 1 (400)、第一の記憶領域 411 などを備えている。第二の記憶装置 40 は、第二の記憶領域 412などを備えている。

【0029】

差分管理テーブル 1 (400) は、第一の記憶領域 411 に書き込んだデータの複製が第二の記憶装置 40 に未だ送信されておらず、データの複製が第二の記憶領域 412 に書き込まれていないことを示す情報である。なお、第二の記憶装置 40 の差分管理テーブル 1 (400) はフェールオーバー時に第一の記憶装置 30 から送信されてくるものである。差分管理テーブル 2 (401) は、第三の記憶領域 413 に書き込んだデータが第一の記憶装置 30 に未だ送信されておらず、データが第一の記憶領域 411 に書き込まれていないことを示す情報である。差分管理テーブル 2 (401) は、第二の記憶装置 40 にあらかじめ持たせることとしてもよいし、フェールオーバー時に作成させるようにしてもよい。なお、第一の記憶装置 30 の差分管理テーブル 2 (401) はフェールバック時に第二の記憶装置 40 から送信されてくるものである。これらの差分管理テーブル 1, 2 (401, 401) により、記憶装置 30, 40 は後述するリモートコピーを行う際に、他の記憶装置 30, 40 に未転送のデータがあるかどうかを判断することができる。また、差分管理テーブル 1, 2 (400, 401) により、記憶装置 30, 40 のチャンネル制御部 101 はどのブロックのデータがどの記憶装置に存在するかを把握することができるようになる。

【0030】

図 6 に差分管理テーブル 1, 2 (400, 401) の一例を示す。差分管理テーブル 1, 2 (400, 401) のビット値の欄には、「1」又は「0」が記録されている。「1」が記録されている場合には、そのブロックには他の記憶装置 30, 40 に未転送のデータがあることを意味する。一方、「0」が記録されている場合には、そのブロックには他の記憶装置 30, 40 に未転送のデータがないことを意味する。なお、セットアップ時には、差分管理テーブル 1, 2 (400, 401) のビット値の欄は全て「0」が記録されている。本実施の形態においては、差分管理テーブル 1, 2 (400, 401) は、共有メモリ 104 に記憶されていることとしているが、記憶領域 411~414 などに記憶されていることとしてもよい。

【0031】

第一の記憶領域 411 は、第一の情報処理装置 10 又は第二の記憶装置 40 から送信されてくるデータの書き込み要求に応じてそのデータを記憶するためのものである。第四の記憶領域 414 は予備の記憶領域である。第二の記憶領域 412 は、第二の情報処理装置 20 又は第一の記憶装置 30 から送信されてくるデータの書き込み要求に応じてそのデータを記憶するためのものである。第三の記憶領域 413 は予備の記憶領域である。なお、本実施の形態においては、記憶領域 411~414 の容量は全て同じサイズに設定されていることとするが、これらに限定されるものではない。

【0032】

次に、記憶装置 30、40 が情報処理装置 10、20 からデータの入出力要求を受信した場合に記憶装置 30、40 において行われる処理を説明する。チャンネル制御部 101 は、情報処理装置 10、20 から送信されてくるデータの書き込み要求を受信すると、データの書き込み要求のコマンド（以下、「データの書き込みコマンド」と称する）を共有メモリ 104 に記憶するとともに、このデータの書き込み要求の対象となるデータ（以下、「書き込みデータ」と称する）をキャッシュメモリ 105 に記憶する。ディスク制御部 103 は、リアルタイムに共有メモリ 104 の内容を監視している。ディスク制御部 103 は、この監視により共有メモリ 104 にデータの書き込みコマンドが書き込まれていることを検知すると、キャッシュメモリ 105 から書き込みデータを読み出し、データの書き込みコマンドに指定されたアドレス（ブロック番号）に基づいて、その書き込みデータを記憶領域 411～414 に書き込む。

【0033】

ディスク制御部 103 は、記憶領域 411～414 にデータの書き込みが完了すると、その旨をチャンネル制御部 101 に通知する。チャンネル制御部 101 は、データの書き込み完了通知を受信すると、情報処理装置 10、20 に対してデータの書き込み完了通知を送信する。

【0034】

チャンネル制御部 101 は、ディスク制御部 103 から前記通知を受領すると、チャンネル制御部 101 は、当該データを書き込んだブロック番号に基づいて、差分管理テーブル 1（400）のビット値の欄を「1」に更新する。なお、ビット値の欄の更新は、ディスク制御部 103 が行うこととしてもよい。そして、チャンネル制御部 101 は、キャッシュメモリ 105 に記憶されている読み出しデータを情報処理装置 10、20 に送信する。

【0035】

一方、チャンネル制御部 101 が、情報処理装置 10、20 から送信されてくるデータの読み出し要求を受信すると、データの読み出し要求のコマンド（以下、「データの読み出しコマンド」と称する）をディスク制御部 103 に送出する。なお、チャンネル制御部 101 からディスク制御部 103 へのデータの読み出しコマンドの伝達は、共有メモリ 104 を介して行われてもよい。

【0036】

ディスク制御部 103 は、チャンネル制御部 101 からデータの読み出しコマンドを受領すると、データの読み出しコマンドにより指定されたアドレスに基づいて、その読み出し対象のデータ（以下、「読み出しデータ」と称する）を記憶領域 411～414 から読み出す。そして、この読み出したデータをキャッシュメモリ 105 に書き込む。ディスク制御部 103 は、キャッシュメモリ 105 へのデータの転送が完了すると、その旨をチャンネル制御部 101 に通知する。

【0037】

なお、上述のように、本実施の形態に係るストレージシステム 90 においては第一の情報処理装置 10 と第一の記憶装置 30 との間でデータの授受を行っているが、バックグラウンドでは、第一の記憶装置 30 のデータの複製を第二の記憶装置 40 においても管理する処理（リモートコピー）が行われている。このリモートコピーにより、対応づけられている記憶領域 411、412 間の内容が一致することとなり、データの冗長管理が可能となる。以下、リモートコピーについて説明する。

【0038】

===リモートコピー===

図 7 は本実施の形態に係る、第一の記憶装置（複製元の記憶装置）30 から第二の記憶装置（複製先の記憶装置）40 へのリモートコピーに関する処理を説明するフローチャートである。

【0039】

第一の記憶装置 30 のチャンネル制御部 101 は、第一の情報処理装置 10 にデータの書

き込み完了通知を送信した後に、共有メモリ 104 に記録されている差分管理テーブル 1 (400) を参照して、第二の記憶装置 40 に未転送のデータの書き込み要求を送信する (S700)。なお、このデータの書き込みコマンドに指定されるアドレスは、送信する書き込みデータが記憶されている第一の記憶領域 411 のブロック番号が記録される。

第二の記憶装置 40 のチャンネル制御部 101 は、データの書き込み要求を第一の記憶装置 30 から受信すると (S701)、ディスク制御部 103 はデータの書き込みコマンドに従って第二の記憶領域 412 に書き込みデータを書き込む (S702)。第二の記憶装置 40 のチャンネル制御部 101 は、第二の記憶領域 412 にデータの書き込みが完了すると、第一の記憶装置 30 にデータの書き込み完了通知を送信する (S703)。

第一の記憶装置 30 のチャンネル制御部 101 は、データの書き込み完了通知を受信すると (S704)、データの書き込み要求に指定されたブロック番号に基づいて、そのブロックに対応する差分管理テーブル 1 (400) のビット値を「1」から「0」に更新する。なお、この更新はディスク制御部 103 が行うこととしてもよい。

次に、第一の記憶装置 30 のチャンネル制御部 101 は、更新した差分管理テーブル 1 (400) を参照して、第二の記憶装置 40 に未転送のデータがあるか否かを判断する (S705)。第一の記憶装置 30 のチャンネル制御部 101 は、第二の記憶装置 40 に未転送のデータが存在すると判断した場合 (S705; YES) には (S700) に進む。一方、第一の記憶装置 30 のチャンネル制御部 101 は、第二の記憶装置 40 に未転送のデータが存在しないと判断した場合 (S705; NO) には処理を終了する。

【0040】

以上のように、第一の記憶装置 30 は第一の記憶領域 411 に記憶したデータの複製を第二の記憶装置 40 に送信し、これを受信した第二の記憶装置 40 は第二の記憶領域 412 に前記データの複製を記憶する処理（以下、「第一の処理」と称する）によって、第一の記憶領域 411 と第二の記憶領域 412 との内容を一致させることが可能となり、データを冗長管理することが可能となる。

【0041】

なお、フェールオーバーした後に第一の処理が終了した場合には、まず、第二の記憶装置 40 のディスク制御部 103 が第三の記憶領域 413 に記憶されているデータの複製を全て第二の記憶領域 412 に書き込む処理が行われる。そして、上述の (S700) ~ (S705) に準じた処理がなされることとなる。なお、(S700) では、第二の記憶装置 40 のチャンネル制御部 101 が、差分管理テーブル 2 (401) を参照して、第一の記憶装置 30 に未転送のデータの書き込み要求を送信することとなる。

【0042】

このように、第二の記憶装置 40 は第三の記憶領域 413 に記憶したデータを第一の記憶装置 30 に送信し、これを受信した第一の記憶装置 30 は第一の記憶領域 411 に前記データを記憶する処理（以下、「第二の処理」と称する）によって、第一の記憶領域 411 と第二の記憶領域 412 との内容を一致させることが可能となる。

【0043】

また、フェールバックした後に第二の処理が終了した場合には、まず、第一の記憶装置 30 のディスク制御部 103 が第四の記憶領域 414 に記憶されているデータの複製を全て第一の記憶領域 411 に書き込む処理が行われる。そして、上述の (S700) ~ (S705) に準じた処理がなされることとなる。なお、(S700) では、第一の記憶装置 30 のチャンネル制御部 101 は、差分管理テーブル 3 を参照して、第二の記憶装置 40 に未転送のデータの書き込み要求を送信することとなる。差分管理テーブル 3 は、第四の記憶領域 414 に書き込んだデータが第二の記憶装置 40 に未だ送信されておらず、データが第二の記憶領域 412 に書き込まれていないことを示す情報である。差分管理テーブル 3 は、第一の記憶装置 30 にあらかじめ持たせることとしてもよいし、フェールバック後のフェールオーバー時に作成させるようにしてもよい。第一の記憶装置 30 の差分管理テーブル 3 はフェールバック後のフェールバック時に第一の記憶装置 30 から第二の記憶装置 40 に送信される。この差分管理テーブル 3 により、第二の記憶装置 40 はリモートコピ

ーを行う際に、第一の記憶装置 30 に未転送のデータがあるかどうかを判断することができる。また、差分管理テーブル 3 により、第一の記憶装置 40 のチャンネル制御部 101 はどのブロックのデータがどの記憶装置に存在するかを把握することができるようになる。なお、差分管理テーブル 3 は、差分管理テーブル 1, 2 (400, 401) と同様の情報が記録されている。

【0044】

このように、第一の記憶装置 30 は第四の記憶領域 414 に記憶したデータを第二の記憶装置 40 に送信し、これを受信した第二の記憶装置 40 は第二の記憶領域 412 に前記データを記憶する処理（以下、「第三の処理」と称する）によって、第一の記憶領域 411 と第二の記憶領域 412 との内容を一致させることが可能となる。

【0045】

== 第一の情報処理装置に障害が発生した場合の処理 ==

次に、図 8 を用いて第二の情報処理装置 20 が第一の情報処理装置 10 の障害を検知した場合の処理の一例を説明する。

第二の情報処理装置 20 が第一の情報処理装置 10 の障害を検知すると (S800)、第二の情報処理装置 20 は第二の記憶装置 40 にフェールオーバーする旨の通知を行う (S801)。

第二の記憶装置 40 のチャンネル制御部 101 は、第二の情報処理装置 20 からフェールオーバーする旨の通知を受信すると (S802)、第一の記憶装置 30 に差分管理テーブル 1 (400) を要求する (S803)。

第一の記憶装置 30 のチャンネル制御部 101 は、第二の記憶装置 40 から差分管理テーブル 1 (400) の要求を受信すると (S804)、共有メモリ 104 に記憶されている差分管理テーブル 1 (400) の複製を第二の記憶装置 40 に送信する (S805)。

第二の記憶装置 40 のチャンネル制御部 101 は、第一の記憶装置 30 から差分管理テーブル 1 (400) を受信すると (S806)、共有メモリ 104 に差分管理テーブル 1 (400) を記憶する (S807)。そして、第二の記憶装置 40 のチャンネル制御部 101 はデータ入出力要求を受け付けることができる旨を第二の情報処理装置 20 に通知する (S808)。

第二の情報処理装置 20 は、第二の記憶装置 40 からデータ入出力要求を受け付けることができる旨の通知を受信すると (S809)、第一の情報処理装置 10 で行っていた処理を引き継ぎ（フェールオーバーする）(S810)、処理を終了する。その後、第二の情報処理装置 20 は第二の記憶装置 40 にデータ入出力要求を送信し、第二の記憶装置 40 は第二の情報処理装置 20 から送信されてくるデータ入出力要求に応じてデータを処理することとなる。

【0046】

なお、本実施の形態においては、第二の情報処理装置 20 が第一の情報処理装置 10 の障害を検知した場合の処理について説明したが、フェールバックする場合にも (S801) ~ (S810) に準じた処理がなされる。なお、フェールバックする旨の通知は、例えば、第一の情報処理装置 10 の障害が復旧した後に行われる。

【0047】

また、本実施の形態においては、フェールオーバーする前に第二の記憶装置 40 が第一の記憶装置 30 から差分管理テーブル 1 (400) を受信することとしているが、フェールオーバー後に第二の記憶装置 40 が第一の記憶装置 30 から差分管理テーブル 1 (400) を受信することとしてもよい。

【0048】

また、本実施例の形態においては、フェールオーバーする旨の通知を受信すると、(S803) ~ (S810) の処理を行うこととしているが、最初のデータ入出力要求を情報処理装置 10, 20 から受信した場合に (S803) ~ (S810) の処理を行うこととしてもよい。

【0049】

また、本実施の形態においては、第二の情報処理装置 20 が第一の情報処理装置 10 の障害を監視することとしているが、第一の記憶装置 30 が第一の情報処理装置 10 の障害を監視することとしてもよい。この場合には、第一の記憶装置 30 が第一の情報処理装置 10 の障害を検知すると、第一の記憶装置 30 は差分管理テーブル 1 (400) を第二の記憶装置 40 に送信し、第二の記憶装置 40 は第二の情報処理装置 20 にデータ入出力要求を受け付けることができる旨を送信することができるようにしてもよい。これにより、第二の情報処理装置 20 はフェールオーバを実行し、データ入出力要求を第二の記憶装置 40 に送信することができるようになる。

【0050】

また、本実施の形態においては、記憶装置 30, 40 が他の記憶装置 30, 40 に差分管理テーブル 1, 2 (400, 401) を要求することにより、記憶装置 30, 40 が受信することとしているが、記憶装置 30, 40 が管理テーブル 1, 2 (400, 401) を更新する度にそれを他の記憶装置 30, 40 に送信することとしてもよい。なお、記憶装置 30, 40 から他の記憶装置 30, 40 に送信される管理テーブル 1, 2 (400, 401) は、情報処理装置 10, 20 を経由して送信されることとしてもよい。また、記憶装置 30, 40 は、上記差分管理テーブル 1, 2 (400, 401) の代わりにデータの書き込み要求で情報処理装置 10, 20 が指定したアドレス (ブロック番号) の情報を他の記憶装置 30, 40 に送信することとしてもよい。

【0051】

以上の仕組みによれば、情報処理装置 10, 20 は、フェールオーバ又はフェールバックする時に行われる処理 (複製先の記憶領域 411, 412 の内容と複製元の記憶領域 412, 411 の内容とを一致させる処理) を待つことにより生じるタイムアウトを回避することができ、情報処理装置 10, 20 間の処理の引き継ぎをスムーズに行うことができる。また、ストレージシステムの運用を円滑に行うことが可能となり、ストレージシステムの信頼性や可用性の向上を図ることが可能となる。

【0052】

===データの書き込み要求を受信した場合の処理===

次に、図 9 を用いてフェールオーバ後に第二の記憶装置 40 が第二の情報処理装置 20 からデータの書き込み要求を受信した場合の処理の一例を説明する。

【0053】

第二の記憶装置 40 のチャンネル制御部 101 は、第二の情報処理装置 20 からデータの書き込み要求を受信すると (S900)、差分管理テーブル 1 (400) を参照して差分管理テーブル 1 (400) のビット値が全て「0」であるか否かを判断する (S901)。第二の記憶装置 40 のチャンネル制御部 101 が差分管理テーブル 1 (400) のビット値が全て「0」ではないと判断した場合 (S901; NO) には、(S904) へ進む。一方、第二の記憶装置 40 のチャンネル制御部 101 が差分管理テーブル 1 (400) のビット値が全て「0」であると判断した場合 (S901; YES) には、(S902) へ進む。

【0054】

(S902) では、第二の記憶装置 40 のチャンネル制御部 101 が、第三の記憶領域 413 に記憶したデータを第二の記憶領域 412 に記憶したか否かを判断する (S903)。第二の記憶装置 40 のチャンネル制御部 101 が第三の記憶領域 413 に記憶した全てのデータの複製を第二の記憶領域 412 に記憶していないと判断した場合 (S902; NO) には、(S904) へ進む。一方、第二の記憶装置 40 のチャンネル制御部 101 が第三の記憶領域 413 に記憶した全てのデータの複製を第二の記憶領域 412 に記憶したと判断した場合 (S902; YES) には、第二の記憶装置 40 のチャンネル制御部 101 は、データの書き込みコマンドを共有メモリ 104 に記憶するとともに、この書き込みデータをキャッシュメモリ 105 に記憶する。ディスク制御部 103 は、共有メモリ 104 にデータの書き込みコマンドが書き込まれていることを検知すると、キャッシュメモリ 105 から書き込みデータを読み出し、データの書き込みコマンドに指定されたブロック番号に

基づいて、第二の記憶領域 412 に書き込みデータを書き込む (S903)。

【0055】

(S904) では、(S903) と同様に第二の記憶装置 40 のディスク制御部 103 が第三の記憶領域 413 に書き込みデータを書き込む。

第二の記憶装置 40 のディスク制御部 103 は、記憶領域 412, 413 にデータの書き込みが完了すると (S903, S904)、その旨を第二の記憶装置 40 のチャネル制御部 101 に通知する。第二の記憶装置 40 のチャネル制御部 101 は、データの書き込み完了通知を受信すると、データの書き込み要求に指定されたブロック番号に基づいて、差分管理テーブル 2 のブロック番号のビット値の欄を「1」に更新する (S905)。そして、第二の記憶装置 40 のチャネル制御部 101 は、第二の情報処理装置 20 にデータの書き込みが完了した旨を通知し (S906)、処理を終了する。

【0056】

なお、本実施の形態においては、(S904) においてデータの書き込み要求の対象となるデータを第三の記憶領域 413 に記憶することとしたが、第二の記憶装置 40 のチャネル制御部 101 は第一の記憶装置 30 にデータの書き込み要求を転送し、第一の記憶装置 30 のチャネル制御部 101 が前記データの書き込み要求に応じてそのデータを第一の記憶領域 411 に記憶することとしてもよい。この場合には、第一の記憶装置 30 のチャネル制御部 101 は、データの書き込みが完了した後に、データの書き込み要求に指定されたブロック番号に基づいて、差分管理テーブル 1 (400) のビット値の欄を「1」に更新することとなる。また、第二の記憶装置 40 のチャネル制御部 101 も、第一の記憶装置 30 からデータの書き込みが完了した旨の通知を受信した場合に、データの書き込み要求に指定されたブロック番号に基づいて、差分管理テーブル 1 (400) のビット値の欄を「1」に更新することとなる。

【0057】

なお、上述したように、第一の処理が終了した場合には、第二の記憶装置 40 のディスク制御部 103 が第三の記憶領域 413 に記憶されているデータの複製を第二の記憶領域 412 に記憶し、その後、第二の処理が開始される。このような手順で処理させることにより、第二の記憶領域 412 に最新のデータを確保させることができ、また、第一の記憶領域 411 においても最新のデータを冗長管理させることが可能となる。

【0058】

本実施の形態においては、フェールオーバーした後の処理について説明したが、フェールバックした後も (S900) ~ (S906) に準じた処理がなされる。この場合において、(S902) の「第三の記憶領域 413」は「第四の記憶領域 414」とし、(S905) の「差分管理テーブル 2 (401)」は「差分管理テーブル 3」とする。差分管理テーブル 3 は、第四の記憶領域 414 に書き込んだデータの複製が第二の記憶装置 40 に未だ送信されておらず、データの複製が第二の記憶領域 412 に書き込まれていないことを示す情報である。差分管理テーブル 3 は、フェールバックする時に第一の記憶装置 30 のチャネル制御部 101 により、新たに作成されて共有メモリ 104 に記憶される。

【0059】

なお、第一の処理の実行中にフェールバックした場合にも、(S900) ~ (S906) に準じた処理がなされる。バックグラウンドでは、上述したように、第一の処理が終了した後に、第二の記憶装置 40 のディスク制御部 103 が第三の記憶領域 413 に記憶されているデータの複製を第二の記憶領域 412 に記憶する。そして、第二の処理が開始され、第二の処理が終了すると、第一の記憶装置 30 のディスク制御部 103 が第四の記憶領域 414 に記憶されているデータの複製を第一の記憶領域 411 に記憶する。その後、第三の処理が開始されることとなる。このような手順で処理することにより、最新のデータを確保することができ、最新のデータを冗長管理することができるようになる。

【0060】

==データの読み出し要求を受信した場合の処理==

次に、図 10 を用いてフェールオーバー後に第二の記憶装置 40 が第二の情報処理装置 2

0 からデータの読み出し要求を受信した場合の処理の一例を説明する。

【0061】

第二の記憶装置 40 のチャネル制御部 101 は、第二の情報処理装置 20 からデータの読み出し要求を受信すると (S1000)、第三の記憶領域 413 にデータの読み出しコマンドに指定されたデータがあるか否かを判断する (S1001)。この判断は、例えば、差分管理テーブル 2 (401) を参照することにより行うことができる。第二の記憶装置 40 のチャネル制御部 101 が第三の記憶領域 413 にデータの読み出しコマンドに指定されたデータがあると判断した場合 (S1001; YES) には、(S1008) へ進む。一方、第二の記憶装置 40 のチャネル制御部 101 が第三の記憶領域 413 にデータの読み出しコマンドに指定されたデータがないと判断した場合 (S1001; NO) には、第二の記憶装置 40 のチャネル制御部 101 は第二の記憶領域 412 にデータの読み出しコマンドに指定されたデータがあるか否かを判断する (S1002)。この判断は、例えば、差分管理テーブル 1 (400) を参照することにより行うことができる。なお、上述の手順で判断することにより、第二の記憶装置 40 は、最新のデータが何処にあるかを判断することが可能となる。

【0062】

(S1002) において、第二の記憶装置 40 のチャネル制御部 101 が第二の記憶領域 412 にデータの読み出しコマンドに指定されたデータがあると判断した場合 (S1002; YES) には、(S1009) へ進む。一方、第二の記憶装置 40 のチャネル制御部 101 が第二の記憶領域 412 にデータの読み出しコマンドに指定されたデータがないと判断した場合 (S1002; NO) には、第二の記憶装置 40 のチャネル制御部 101 は、第一の記憶装置 30 にデータの読み出し要求を送信する (S1003)。

【0063】

第一の記憶装置 30 のチャネル制御部 101 は第二の記憶装置 40 からデータの読み出し要求を受信すると (S1004)、データの読み出しコマンドをディスク制御部 103 に送出する。ディスク制御部 103 は、チャネル制御部 101 からデータの読み出しコマンドを受領すると、データの読み出しコマンドにより指定されたブロック番号に基づいて、その読み出しデータを第一の記憶領域 411 から読み出す (S1005)。そして、この読み出したデータをキャッシュメモリ 105 に書き込む。ディスク制御部 103 は、キャッシュメモリ 105 へのデータの転送が完了すると、その旨をチャネル制御部 101 に通知する。第二の記憶装置 40 のチャネル制御部 101 は、ディスク制御部 103 から前記通知を受領すると、キャッシュメモリ 105 に記憶されている読み出しデータを第二の記憶装置 40 に送信する (S1006)。

【0064】

第二の記憶装置 40 のチャネル制御部 101 は、第一の記憶装置 30 から読み出しデータを受信すると (S1007)、第二の情報処理装置 20 に読み出しデータを送信し (S1010)、処理を終了する。

【0065】

(S1008) では、ディスク制御部 103 が第二の記憶装置 30 のチャネル制御部 101 からデータの読み出しコマンドを受領すると、ディスク制御部 103 はデータの読み出しコマンドにより指定されたブロック番号に基づいて、第三の記憶領域 413 から読み出しデータを読み出す。そして、第二の記憶装置 40 のチャネル制御部 101 は第二の情報処理装置 20 に読み出しデータを送信し (S1010)、処理を終了する。

【0066】

(S1009) では、(S1008) と同様にディスク制御部 103 が第二の記憶領域 412 から読み出しデータを読み出す。そして、第二の記憶装置 40 のチャネル制御部 101 は第二の情報処理装置 20 に読み出しデータを送信し (S1010)、処理を終了する。

【0067】

本実施の形態においては、フェールオーバーした後の処理について説明したが、フェール

バックした後も (S1000) ~ (S1006) に準じた処理がなされる。この場合において、(S1001) では、第一の記憶装置 30 のチャネル制御部 101 が、第四の記憶領域 414 にデータの読み出しコマンドに指定されたデータがあるか否かを判断する。この判断は、例えば、差分管理テーブル 3 を参照することにより行うことができる。また、(S1002) では、第一の記憶装置 30 のチャネル制御部 101 が、第三の記憶領域 413 にデータの読み出しコマンドに指定されたデータがあるか否かを判断する。この判断は、例えば、差分管理テーブル 3 を参照することにより行うことができる。

なお、第三の記憶領域 413 にデータの読み出しコマンドに指定されたデータがあると判断した場合 (S1002; YES) には、第一の記憶装置 30 のチャネル制御部 101 は第二の記憶装置 40 にデータの読み出し要求を送信することとなる (S1003)。一方、第三の記憶領域 413 にデータの読み出しコマンドに指定されたデータがないと判断した場合 (S1002; NO) には、第一の記憶装置 30 のチャネル制御部 101 は第一の記憶領域 411 からデータを読み出すこととなる (S1009)。なお、第一の処理の実行中にフェールバックした場合にも同様の処理が行われることとなる。

【0068】

以上のような手順で判断することにより、記憶装置 30, 40 は、最新のデータが何処にあるかを判断することが可能となる。また、上述のように処理することにより、記憶装置 30, 40 は、情報処理装置 10, 20 から送信されてくるデータの読み出し要求に応じて最新のデータを提供することが可能となる。

【0069】

===その他の実施の形態===

図 11 は他の実施の形態の一例として説明する、ストレージシステム 90 の概略構成を示す図である。

【0070】

記憶装置 30, 40 は、データ位置判定テーブル 500, 600、ローカル未反映テーブル 501, 601、リモート未反映テーブル 502, 602、通常ボリューム 510, 610、ローカル未反映ボリューム 511, 611、リモート未反映ボリューム 512, 612などを備えている。

【0071】

通常ボリューム 510, 610 は、他の記憶装置 30, 40 又は情報処理装置 10, 20 から送信されてくる書き込みデータを書き込むための記憶領域である。ローカル未反映ボリューム 511, 611 は、予備の記憶領域である。情報処理装置 10, 20 から送信されてくる書き込みデータをローカル未反映ボリューム 511, 611 に書き込むと、記憶装置 30, 40 のチャネル制御部 101 又はディスク制御部 103 は、データを書き込んだブロック番号に基づいて、共有メモリ 104 などに記憶されているローカル未反映テーブル 501, 601 の内容を更新する。

【0072】

図 12 にローカル未反映テーブル 501, 601 の一例を示す。ローカル未反映テーブル 501, 601 の各ブロック番号には、通常ボリュームブロック番号の欄、時刻の欄、及び後方ポインタの欄が設けられている。通常ボリュームブロック番号の欄には、情報処理装置 10, 20 から送信されてくるデータの書き込みコマンドに指定されたブロック番号が記録される。なお、初期状態においては、通常ボリュームブロック番号の欄には空きエントリを意味する「-1」が記録されている。この通常ボリュームブロック番号の欄を設けることにより、記憶装置 30, 40 のチャネル制御部 101 やディスク制御部 103 は、ローカル未反映ボリューム 511, 611 の各ブロック番号に記憶されているデータが通常ボリューム 510, 610 のどのブロック番号に記憶しなければならないかを把握することが可能となる。

【0073】

時刻の欄には、情報処理装置 10, 20 から送信されてくるデータの書き込みコマンドのヘッダに記録されている時刻 (例えば、年月日時分秒) が記録される。なお、この時刻

は、例えば、情報処理装置 10, 20 がデータの書き込み要求を作成した時刻であってもよいし、情報処理装置 10, 20 が記憶装置 30, 40 に対してデータの書き込み要求を送信した時刻であってもよい。この時刻は、情報処理装置 10, 20 のタイマー 200 によって監視されており、データの書き込み要求を送信する際にデータの書き込みコマンドのヘッダに記録される。なお、記憶装置 30, 40 のチャンネル制御部 101 が情報処理装置 10, 20 からデータの書き込み要求を受信した時刻を時刻の欄に記録することとしてもよい。この時刻の欄を設けることにより、記憶装置 30, 40 のチャンネル制御部 101 は情報処理装置 10, 20 から受信した書き込みデータを時系列で管理することが可能となる。

【0074】

後方ポインタの欄には、ローカル未反映ボリューム 511, 611 のブロック番号に記憶されているデータの次に情報処理装置 10, 20 から受信した書き込みデータが記憶されているローカル未反映ボリューム 511, 611 のブロック番号が記録されている。この記録は、記憶装置 30, 40 のチャンネル制御部 101 がローカル未反映テーブル 501, 601 の時刻の欄を確認することにより行われる。なお、ローカル未反映ボリューム 511, 611 のあるブロック番号に記憶されているデータより最新のデータが、ローカル未反映ボリューム 511, 611 に記憶されていない場合には、そのデータが最新であること意味する「-1」が記録されることとなる。この後方ポインタの欄を設けることにより、記憶装置 30, 40 のディスク制御部 103 は、リモートコピーを実行する前にローカル未反映ボリューム 511, 611 に記憶されているデータを古いものから順に通常ボリューム 510, 610 に記憶することが可能となり、データの整合性を確保することができるようになる。

【0075】

リモート未反映ボリューム 512, 612 は、記憶装置 30, 40 の通常ボリューム 510, 610 間の内容を一致させるために、他の記憶装置 30, 40 に送信するためのデータが記憶されている記憶領域である。

【0076】

記憶装置 30, 40 が情報処理装置 10, 20 から送信されてくる書き込みデータを、通常ボリューム 510, 610 又はローカル未反映ボリューム 511, 611 に書き込むと、記憶装置 30, 40 はその書き込みデータをリモート未反映ボリューム 512, 612 にも書き込む。そして、記憶装置 30, 40 のチャンネル制御部 101 は、データを書き込んだブロック番号に基づいて、共有メモリ 104 などに格納されているリモート未反映テーブル 502, 602 の内容を更新する。なお、この更新は、記憶装置 30, 40 のディスク制御部 103 が行うこととしてもよい。

【0077】

図 13 にリモート未反映テーブル 502, 602 の一例を示す。リモート未反映テーブル 502, 602 の各ブロック番号には、通常ボリュームブロック番号の欄、時刻の欄、及び後方ポインタの欄が設けられており、それらの欄にはローカル未反映テーブル 501, 601 の場合と同様の内容が記録されている。記憶装置 30, 40 のチャンネル制御部 101 は、リモート未反映テーブル 502, 602 を参照することにより、リモートコピーにおいてリモート未反映ボリューム 512, 612 に記憶されているデータを古いものから順に他の記憶装置 30, 40 に送信することが可能となる。また、他の記憶装置 30, 40 は、古いものから順に送信されてくる書き込みデータを通常ボリューム 510, 610 に記憶することができるようになるので、データの整合性を確保することができるようになる。

【0078】

データ位置判定テーブル 500, 600 は、通常ボリューム 510, 610 に書き込んだデータの複製（リモート未反映ボリューム 512, 612 に書き込んだデータ）が、他の記憶装置 30, 40 に未だ送信されておらず、データの複製が他の記憶装置 30, 40 の通常ボリューム 510, 610 に書き込まれていないことを示す情報である。データ位

置判定テーブル 500, 600 は、他の記憶装置 30, 40 から送信要求があった場合に記憶装置 30, 40 のチャンネル制御部 101 によって作成され、記憶装置 30, 40 から他の記憶装置 30, 40 に送信される。データ位置判定テーブル 500, 600 は、リモート未反映テーブル 502, 602 に基づいて作成される。なお、データ位置判定テーブル 500, 600 の送信要求は、例えば、記憶装置 30, 40 にアクセスする情報処理装置 10, 20 からフェールオーバー又はフェールバックする旨の通知があった場合に行われる。

【0079】

図 14 にデータ位置判定テーブル 500, 600 の一例を示す。リモート未反映テーブル 502, 602 の通常ボリュームブロック番号に対応する欄にブロック番号が記録されている場合には、記憶装置 30, 40 のチャンネル制御部 101 はそのブロック番号に基づいて、データ位置判定テーブル 500, 600 の通常ボリュームブロック番号に対応する欄に「-2」を記録する。また、記憶装置 30, 40 のチャンネル制御部 101 は、それ以外の欄に「-1」を記録する。

【0080】

このようにして作成されたデータ位置判定テーブル 500, 600 を受信した記憶装置 30, 40 のチャンネル制御部 101 は、情報処理装置 10, 20 から送信されてくるデータの読み出し要求の対象となる最新のデータがどのボリュームに存在するのかを把握することができる。

【0081】

また、本実施の形態においては、データ位置判定テーブル 500, 600 を受信した記憶装置 30, 40 は、情報処理装置 10, 20 から送信されてくるデータの書き込み要求に応じたデータをローカル未反映ボリューム 511, 611 に書き込んだ場合には、データの書き込みコマンドに指定されたブロック番号に基づいて、データ位置判定テーブル 500, 600 の通常ボリュームブロック番号に対応する欄にデータを書き込んだブロックの番号を記録する。これにより、記憶装置 30, 40 のチャンネル制御部 101 は、情報処理装置 10, 20 から送信されてくるデータの読み出し要求の対象となるデータがどのボリュームに記憶されているかを把握することができるようになる。なお、記憶装置 30, 40 はデータ位置判定テーブル 500, 600 を受信すると、それを共有メモリ 104 に記憶することとしてもよいし、別途設けられた記憶領域（メモリやボリュームなど）に記憶することとしてもよい。

【0082】

==リモートコピー==

図 15 は他の実施の形態の一例として説明する、バックグラウンドにおいて行われるリモートコピーのフローチャートを示す図である。

【0083】

第一の記憶装置 30 のチャンネル制御部 101 は、第一の情報処理装置 10 にデータの書き込み完了通知を送信した後に、共有メモリ 104 に記憶されているリモート未反映テーブル 502 の時刻の欄を参照して、リモート未反映ボリューム 512 に記憶されている一番古いデータの書き込み要求を作成し、第二の記憶装置 40 に送信する（S1500）。なお、データの書き込みコマンドのアドレスには、リモート未反映テーブル 502 の通常ボリュームブロック番号の欄に記録されているブロック番号が記録されることとなる。そして、（S1501）～（S1504）の処理がなされる。なお、（S1501）～（S1504）の処理は（S701）～（S704）と同様に行われるので、ここでは省略することとする。

【0084】

その後、第一の記憶装置 30 のチャンネル制御部 101 は、リモート未反映テーブル 502 の後方ポインタ欄を確認し、後方ポインタが「-1」であるか否かを判断する（S1505）。後方ポインタが「-1」でないと判断した場合（S1505; NO）には、（S1500）に進み、後方ポインタ欄に記録されているリモート未反映ボリューム 512 の

ブロック番号に基づいて、そのブロックに格納されているデータ（一番古いデータ）の書き込み要求を第二の記憶装置 40 に送信する（S1500）。一方、後方ポインタが「-1」であると判断した場合（S1505; YES）には、第一の記憶装置 30 のチャンネル制御部 101 は、第二の記憶装置 40 に未転送のデータがないと判断して処理を終了させる。

【0085】

なお、（S1505）の判断をした後に、第一の記憶装置 30 のチャンネル制御部 101 は、第二の記憶装置 40 に送信したデータが記憶されていたブロックの番号に基づいて、リモート未反映テーブル 502 の通常ボリュームブロック番号の欄を「-1」に更新する。もし、第二の記憶装置 40 が第一の記憶装置 30 から受信したデータ位置判定テーブル 600 を共有メモリ 104 などに記憶している場合には、（S1502）の処理を行った後に第二の記憶装置 40 のチャンネル制御部 101 は、データを書き込んだブロックの番号に基づいて、データ位置判定テーブル 600 の通常ボリュームブロック番号に対応する欄の数値を削除する。

【0086】

本実施例においては、第一の記憶装置 30 のチャンネル制御部 101 がリモート未反映ボリューム 512 に記憶されているデータを古いものから順に第二の記憶装置 40 に送信するために、リモート未反映テーブル 502 の時刻欄及び後方ポインタ欄を参照することとしているが、第二の記憶装置 40 に送信するデータを時刻の古いものから順に並べて登録するためのキューを設けることとしてもよい。

【0087】

以上のように、第一の記憶装置 30 はリモート未反映ボリューム 512 に記憶したデータを第二の記憶装置 40 に送信し、これを受信した第二の記憶装置 40 は通常ボリューム 610 に前記データを記憶する処理によって、通常ボリューム 510, 610 間の内容を一致させることが可能となる。

【0088】

なお、フェールオーバーした後に、第二の記憶装置 40 の通常ボリューム 610 の内容を第一の記憶装置 30 の通常ボリューム 510 の内容と一致させる処理が終了した場合には、第二の記憶装置 40 のディスク制御部 103 は、ローカル未反映ボリューム 611 に記憶されているデータを通常ボリューム 610 に記憶する。そして、上述の（S1500）～（S1505）に準じた処理が行われることとなる。このように、第二の記憶装置 40 はリモート未反映ボリューム 612 に記憶したデータを第一の記憶装置 30 に送信し、これを受信した第一の記憶装置 30 は通常ボリューム 510 に前記データを記憶する処理によって、通常ボリューム 510, 610 間の内容を一致させることが可能となる。

【0089】

また、フェールバックした後に、第一の記憶装置 30 の通常ボリューム 510 の内容を第二の記憶装置 40 の通常ボリューム 610 の内容と一致させる処理が終了した場合にも、第一の記憶装置 30 のディスク制御部 103 がローカル未反映ボリューム 511 に記憶されているデータを通常ボリューム 510 に記憶してから、（S1500）～（S1505）のフローと同様の処理が行われることになる。

【0090】

なお、記憶装置 30, 40 のディスク制御部 103 がローカル未反映ボリューム 511, 611 に記憶されているデータを通常ボリューム 510, 610 に記憶する処理は、ローカル未反映テーブル 601 の時刻欄及び後方ポインタ欄に基づいて、古いデータから順に行われる。この場合において、通常ボリューム 510, 610 に記憶するデータを古いものから順に並べて登録するためのキューを設けることとしてもよい。

【0091】

=== 第一の情報処理装置に障害が発生した場合の処理 ===

第二の情報処理装置 20 が第一の情報処理装置 10 の障害を検知した場合には、図 8 に示される（S800）～（S810）のフローに準じた処理がなされる。

なお、(S803)では、第二の記憶装置40のチャンネル制御部101は、第一の記憶装置30に対してデータ位置判定テーブル500を要求することとなる。

また、(S805)では、第一の記憶装置30は、リモート未反映テーブル502に基づいてデータ位置判定テーブル500を作成し、第二の記憶装置40に送信することとなる。

なお、本実施の形態においては、第二の情報処理装置20が第一の情報処理装置10の障害を検知した場合の処理について説明したが、フェールバックする場合にも(S801)～(S810)に準じた処理がなされる。

【0092】

===データの書き込み要求を受信した場合の処理===

次に、図16を用いてフェールオーバー又はフェールバックした後に、記憶装置30、40が情報処理装置10、20からデータの書き込み要求を受信した場合の処理の一例を説明する。

【0093】

記憶装置30、40のチャンネル制御部101は、情報処理装置10、20からデータの書き込み要求を受信すると(S1600)、データ位置判定テーブル500、600を参照して、通常ボリュームブロック番号に対応する欄にローカル未反映ボリューム511、611のブロック番号が記録されているかどうかを判断する(S1601)。ローカル未反映ボリューム511、611のブロック番号が記録されていると判断した場合(S1601;YES)には、(S1603)へ進む。一方、ローカル未反映ボリューム511、611のブロック番号が記録されていないと判断した場合(S1601;NO)には、(S1602)へ進む。

【0094】

(S1602)では、記憶装置30、40のチャンネル制御部101は、データの書き込みコマンドを共有メモリ104に記憶するとともに、この書き込みデータをキャッシュメモリ105に記憶する。ディスク制御部103は、共有メモリ104にデータの書き込みコマンドが書き込まれていることを検知すると、キャッシュメモリ105から書き込みデータを読み出し、データの書き込みコマンドに指定されたブロック番号に基づいて、通常ボリューム510、610に書き込みデータを書き込む。

【0095】

(S1603)では、(S1602)と同様に、記憶装置30、40のディスク制御部103がローカル未反映ボリューム511、611の空き領域に書き込みデータを書き込む。その後、チャンネル制御部101がディスク制御部103からデータの書き込み完了通知を受信すると、チャンネル制御部101はローカル未反映テーブル501、601の内容を更新する(S1604)。この更新は、データを書き込んだブロックの番号に基づいて行われる。ローカル未反映テーブル501、601の通常ボリュームブロック番号の欄には、データの書き込みコマンドに指定されたアドレスが記録される。また、ローカル未反映テーブル501、601の時刻の欄には、データの書き込みコマンドのヘッダに記録されている時刻が記録される。その後、チャンネル制御部101は、ローカル未反映テーブル501、601の時刻の欄を参照しながら、ローカル未反映ボリューム511、611に記憶されているデータが古いものから順に通常ボリューム510、610に記憶されるようにするため、後方ポインタの欄を更新する。

【0096】

(S1605)では、記憶装置30、40のディスク制御部103が通常ボリューム510、610又はローカル未反映ボリューム511、611に書き込んだデータをリモート未反映ボリューム512、612に書き込む。その後、チャンネル制御部101がディスク制御部103からデータの書き込み完了通知を受信すると、(S1604)と同様に、チャンネル制御部101はリモート未反映テーブル502、602の内容を更新する(S1606)。そして、記憶装置30、40のチャンネル制御部101は、情報処理装置10、20にデータの書き込みが完了した旨を通知し(S1607)、処理を終了する。

【0097】

===データの読み出し要求を受信した場合の処理===

フェールオーバ又はフェールバックした後に、記憶装置30、40のチャンネル制御部101が、情報処理装置10、20からデータの読み出し要求を受信した場合にも、図10に示される(S1000)~(S1010)と同様の処理がなされる。

【0098】

なお、(S1001)では、記憶装置30、40のチャンネル制御部101が、情報処理装置10、20から受信したデータの読み出しコマンドに指定された最新のデータがローカル未反映ボリューム511、611にあるかどうかを判断する。また、(S1002)では、記憶装置30、40のチャンネル制御部101が、データの読み出しコマンドに指定された最新のデータが通常ボリューム510、610にあるかどうかを判断する。これらの判断は、データの読み出しコマンドに指定されたアドレスに基づいて、データ位置判定テーブル500、600の通常ボリュームブロック番号に対応する欄の値を参照して行われる。

前記欄にローカル未反映ボリューム511、611のブロック番号が記録されている場合には、記憶装置30、40のチャンネル制御部101は、最新のデータがローカル未反映ボリューム511、611にあると判断する(S1001;YES)。また、前記欄に「-1」が記録されている場合には、記憶装置30、40のチャンネル制御部101は、最新のデータが通常ボリューム510、610にあると判断する(S1002;YES)。なお、前記欄に「-2」が記録されている場合には、記憶装置30、40のチャンネル制御部101は、最新のデータが他の記憶装置30、40のリモート未反映ボリューム512、612にあると判断するので、(S1005)では、他の記憶装置30、40がデータの読み出し要求の対象となる最新データをリモート未反映ボリューム512、612から読み出すこととなる。

【0099】

以上、本実施の形態について説明したが、上記実施例は本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に、本発明にはその等価物も含まれる。

【図面の簡単な説明】

【0100】

【図1】本実施の形態に係る、ストレージシステム90の全体構成を示すブロック図である。

【図2】本実施の形態に係る、記憶装置30、40の一例として説明するディスクアレイ装置の具体的な構成を示す図である。

【図3】本実施の形態に係る情報処理装置10、20の構成の一例を示すブロック図である。

【図4】本実施の形態に係るチャンネル制御部101の構成の一例を示すブロック図である。

【図5】本発明の一実施例として説明するストレージシステム90の概略構成を示す図である。

【図6】本実施の形態に係る、差分管理テーブル1、2(400、401)の一例を示す図である。

【図7】本実施の形態に係る、第一の記憶装置30から第二の記憶装置40へのリモートコピーに関する処理を説明するフローチャートである。

【図8】本実施の形態に係る、第二の情報処理装置20が第一の情報処理装置10の障害を検知した場合の処理の一例を示す図である。

【図9】本実施の形態に係る、フェールオーバ後に第二の記憶装置40が第二の情報処理装置20からデータの書き込み要求を受信した場合の処理の一例を示すフローチャートである。

【図10】本実施の形態に係る、フェールオーバ後に第二の記憶装置40が第二の情

報処理装置 20 からデータの読み出し要求を受信した場合の処理の一例を示すフローチャートである。

【図 11】他の実施の形態の一例として説明する、ストレージシステム 90 の概略構成を示す図である。

【図 12】他の実施の形態に係る、ローカル未反映テーブル 501, 601 の一例を示す図である。

【図 13】他の実施の形態に係る、リモート未反映テーブル 502, 602 の一例を示す図である。

【図 14】他の実施の形態に係る、データ位置判定テーブル 500, 600 の一例を示す図である。

【図 15】他の実施の形態の一例として説明する、バックグラウンドにおいて行われるリモートコピーのフローチャートを示す図である。

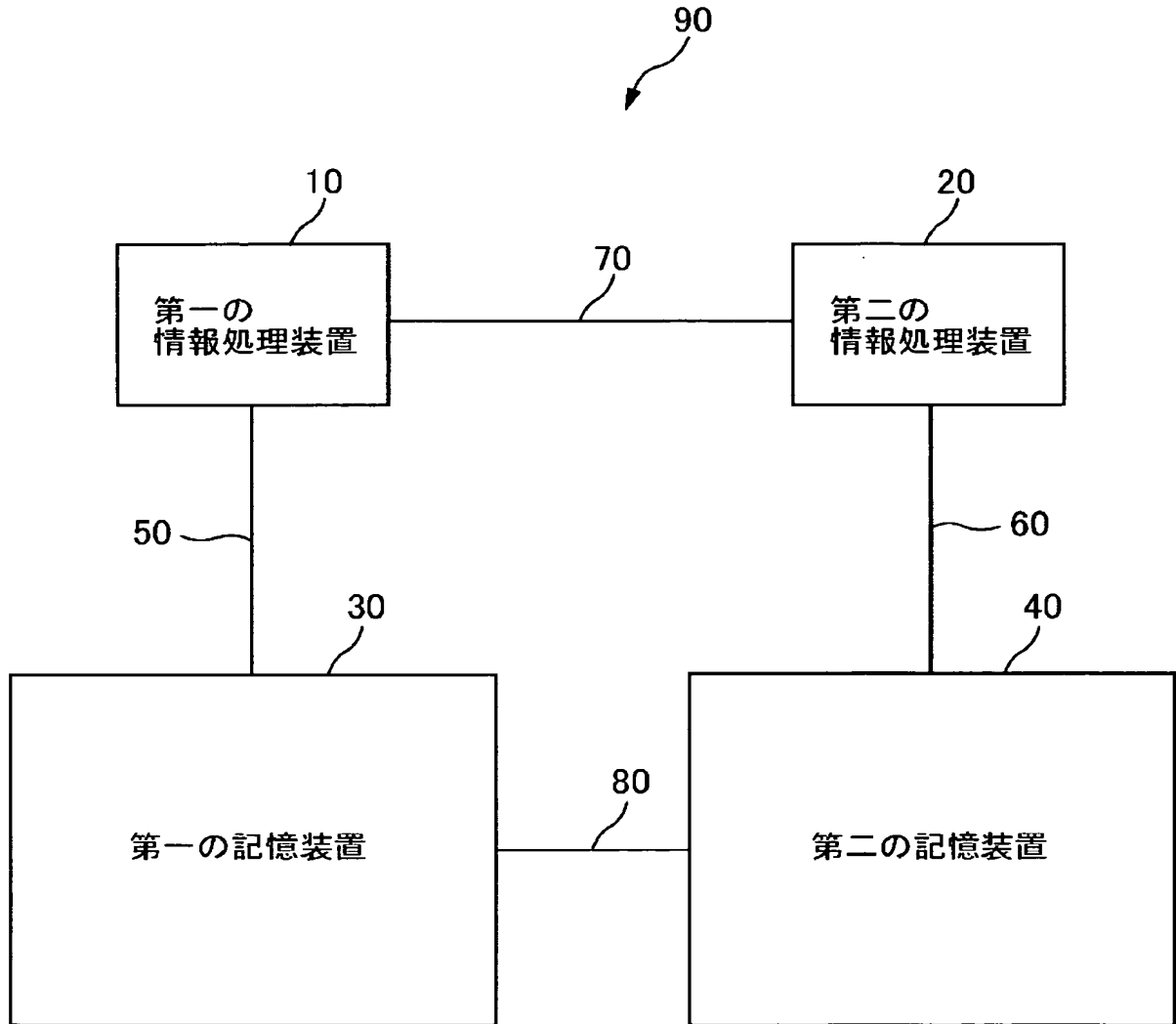
【図 16】他の実施の形態に係る、フェールオーバー後に第二の記憶装置 40 が第二の情報処理装置 20 からデータの書き込み要求を受信した場合の処理の一例を示すフローチャートである。

【符号の説明】

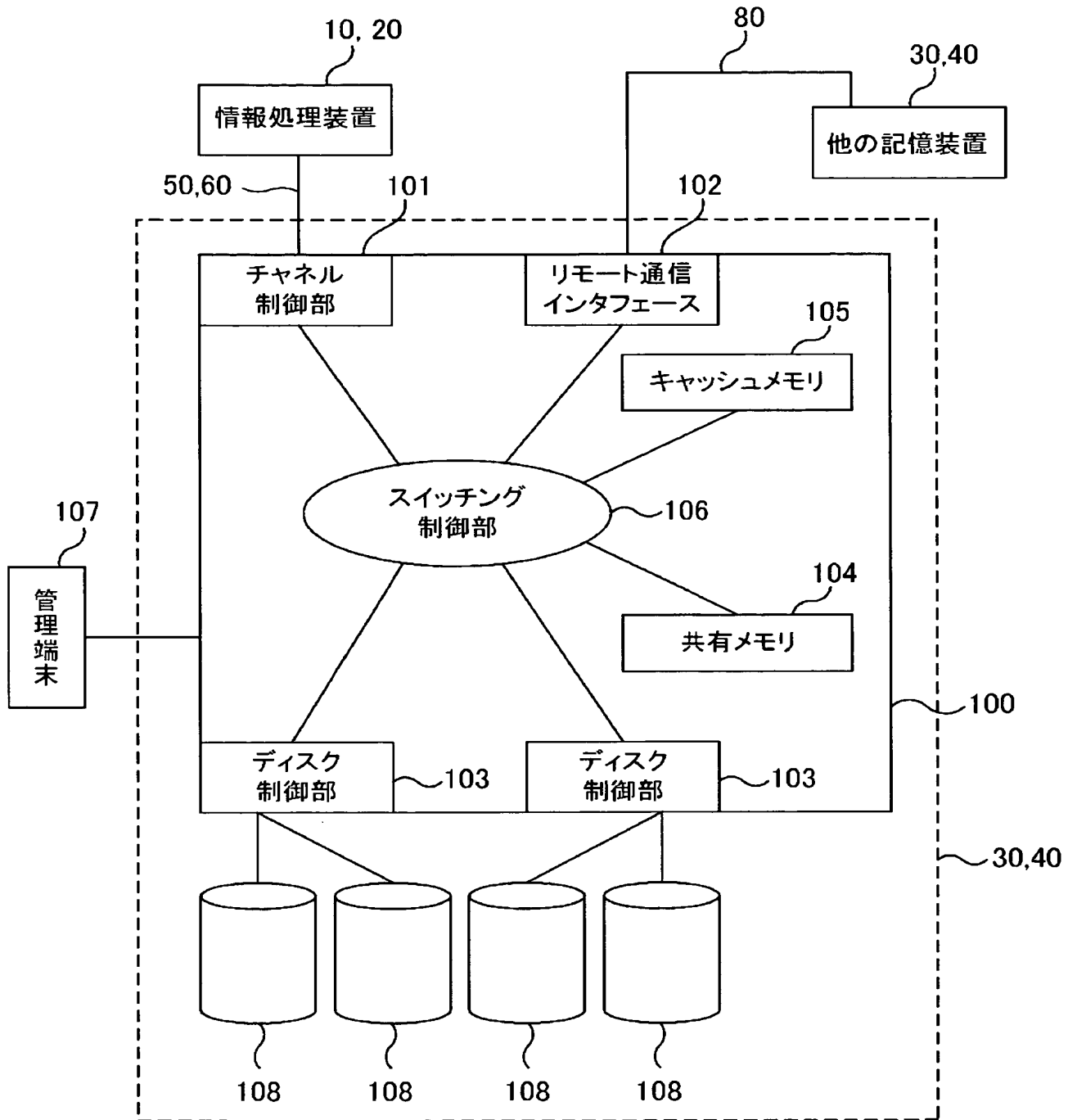
【0101】

10	第一の情報処理装置	20	第二の情報処理装置
30	第一の記憶装置	40	第二の記憶装置
50, 60	第二のネットワーク	70	第三のネットワーク
80	第一のネットワーク	100	記憶デバイス制御装置
101	チャネル制御部	102	リモート通信インタフェース
103	ディスク制御部	104	共有メモリ
105	キャッシュメモリ	106	スイッチング制御部
107	管理端末	108	ディスクドライブ
110	CPU	120	メモリ
130	ポート	140	記憶装置
150	バス	160	記録媒体読取装置
170	入力装置	180	出力装置
190	記録媒体	211	CPU
212	キャッシュメモリ	213	制御メモリ
214	制御プログラム	215	ポート

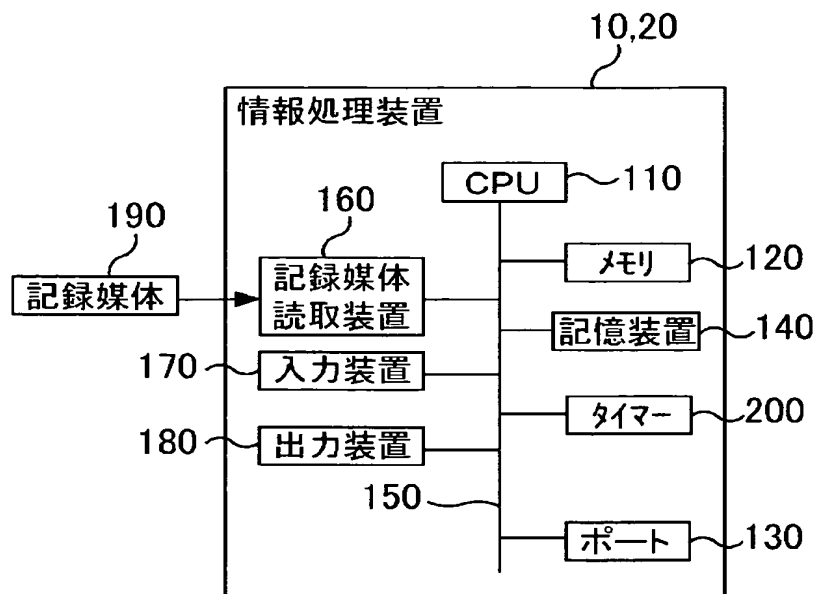
【書類名】 図面
【図 1】



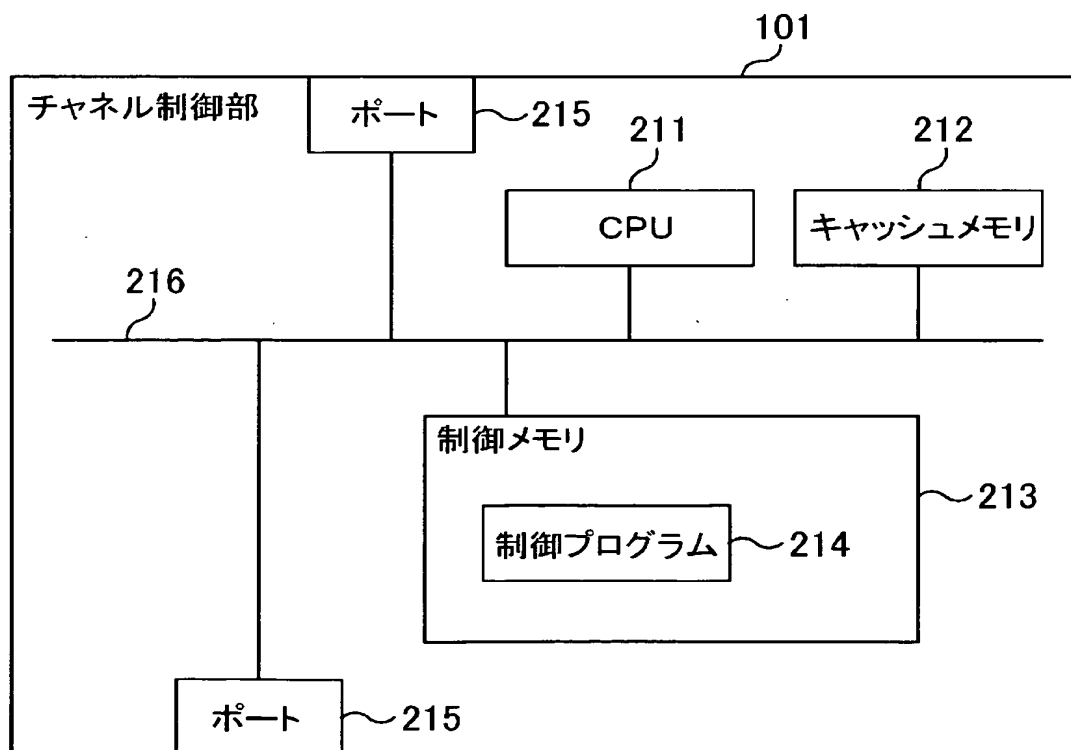
【図 2】



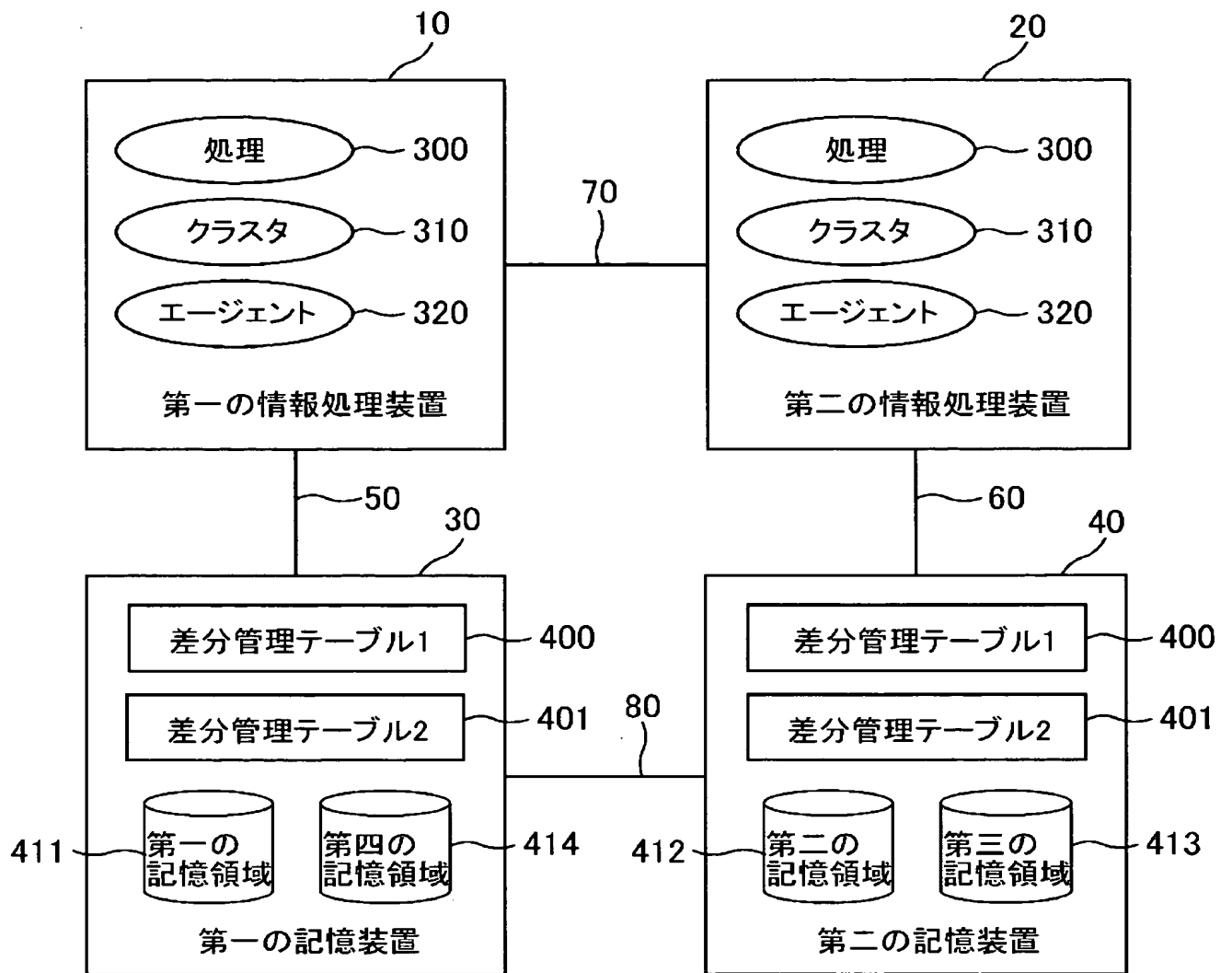
【図 3】



【図 4】



【図 5】

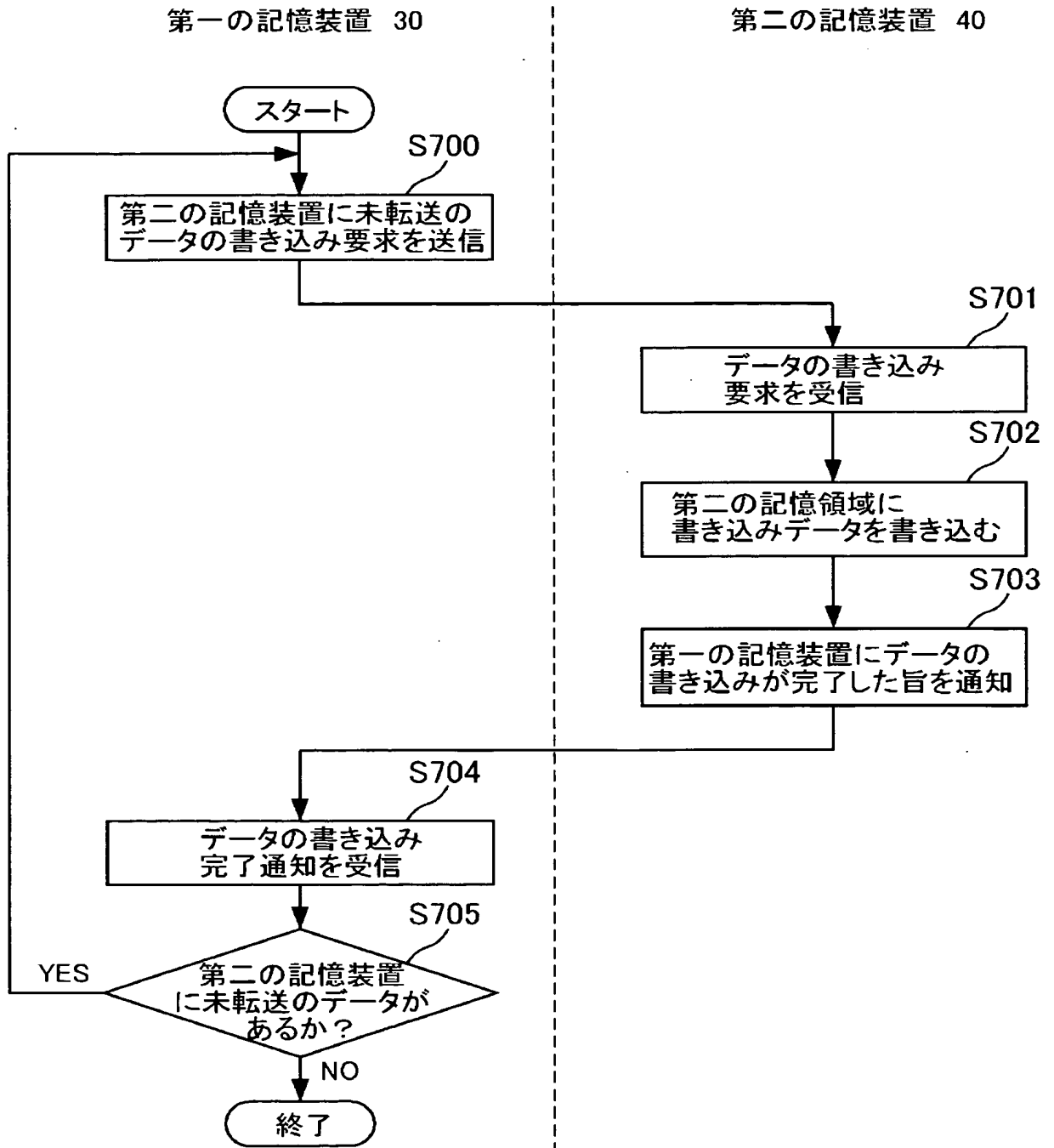


【図 6】

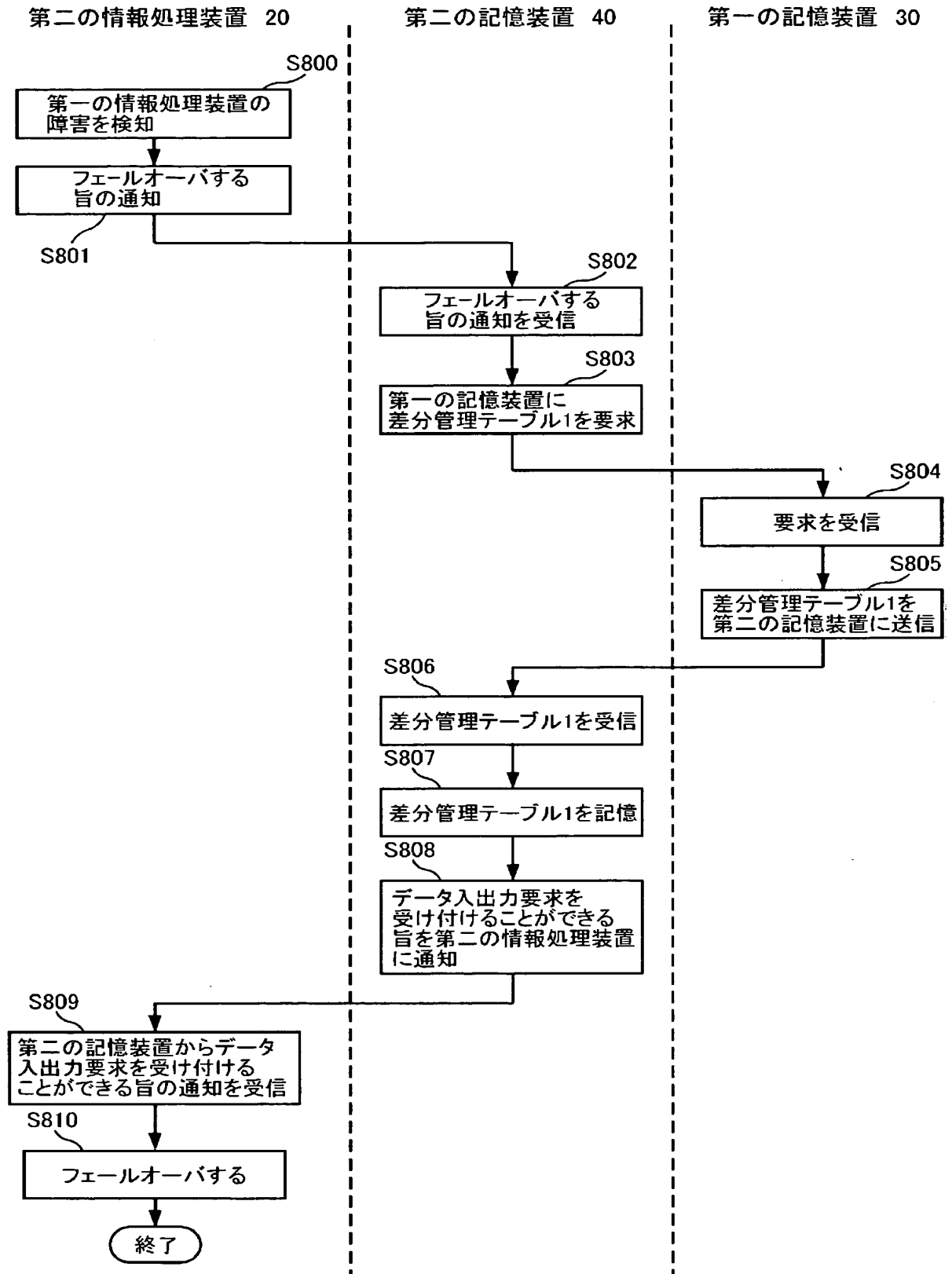
差分管理テーブル 400,401

ブロック番号	ビット値
0	1
1	1
2	0
3	0
⋮	⋮
n	1
⋮	⋮

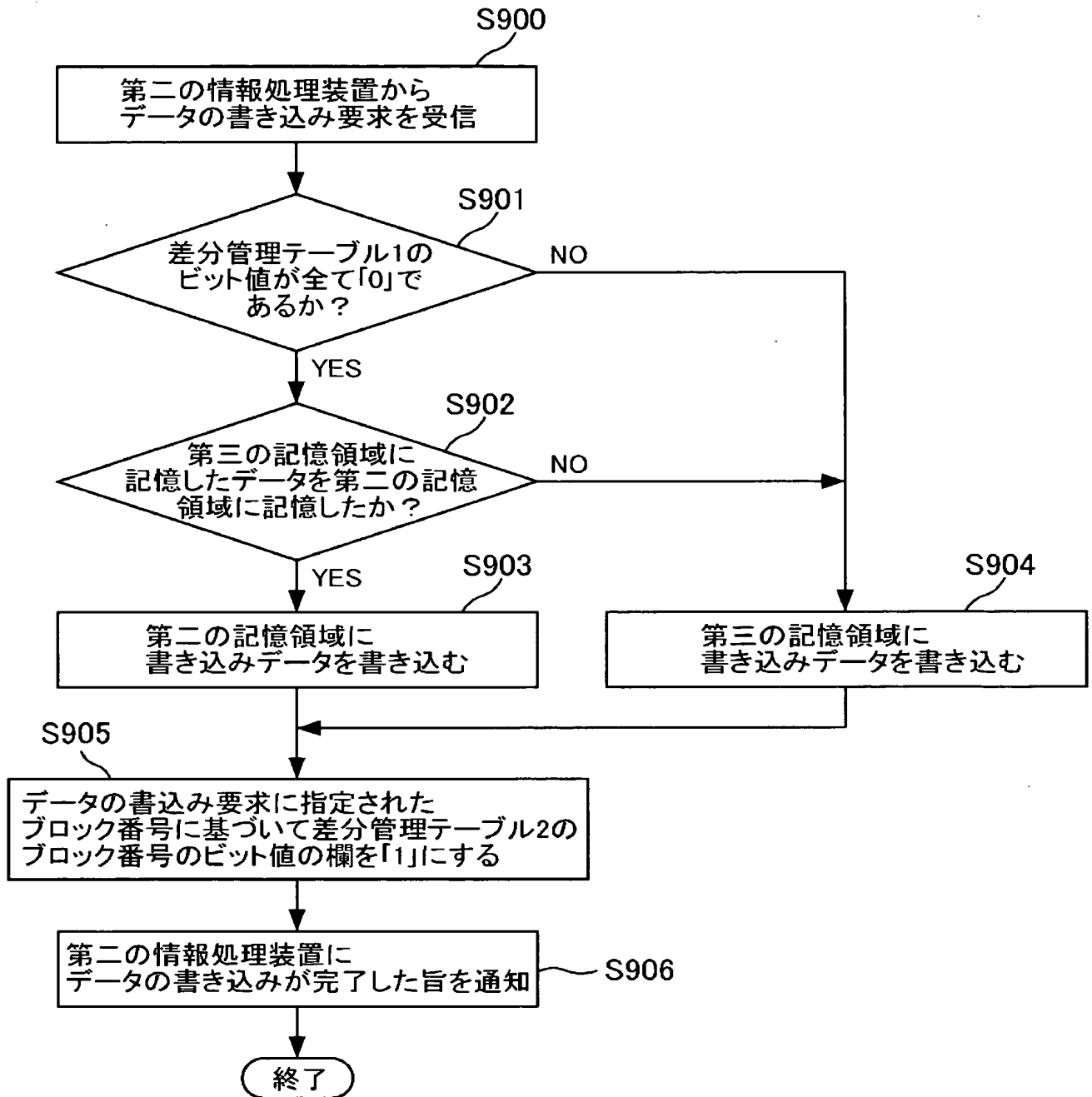
【図 7】



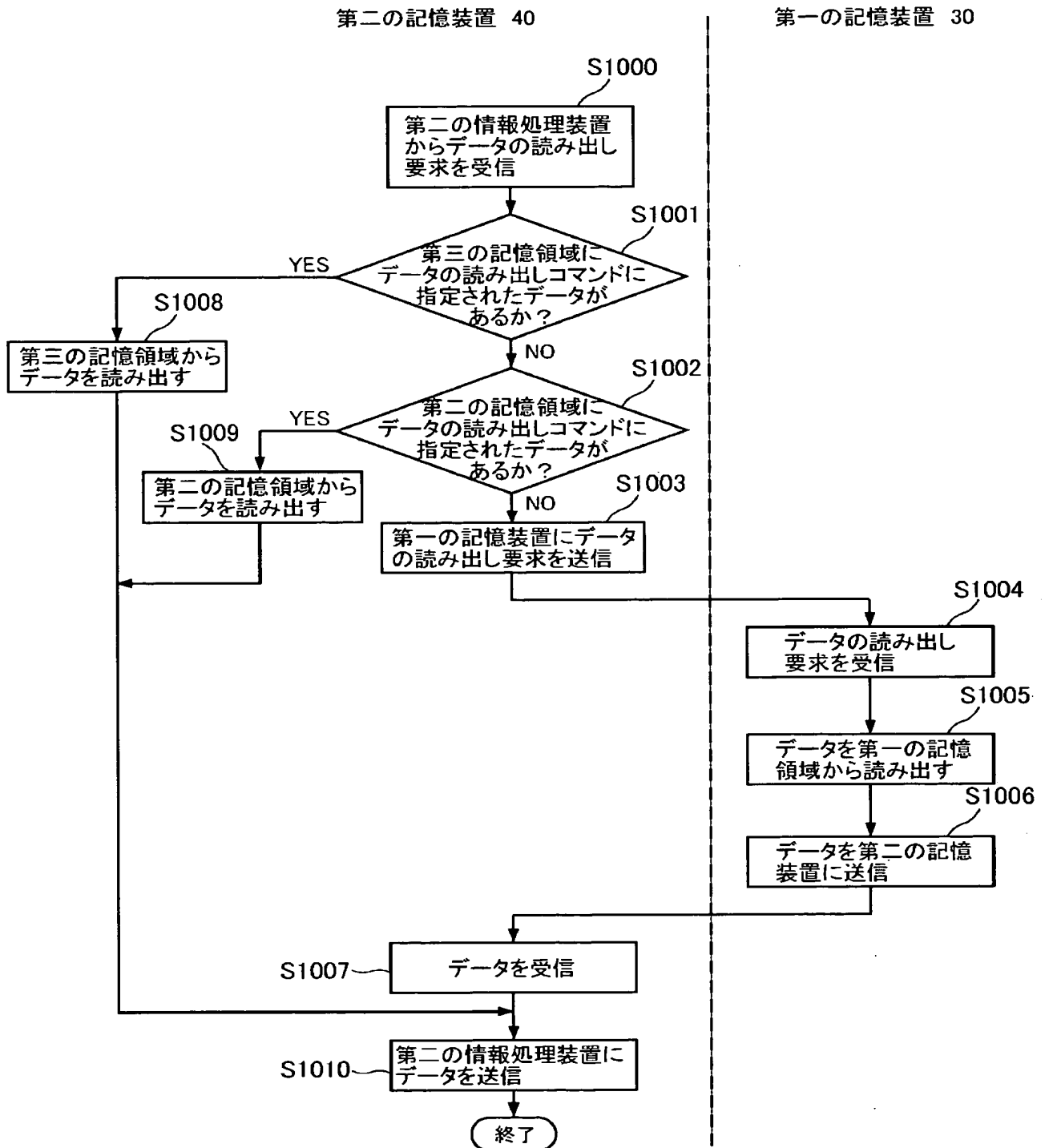
【図 8】



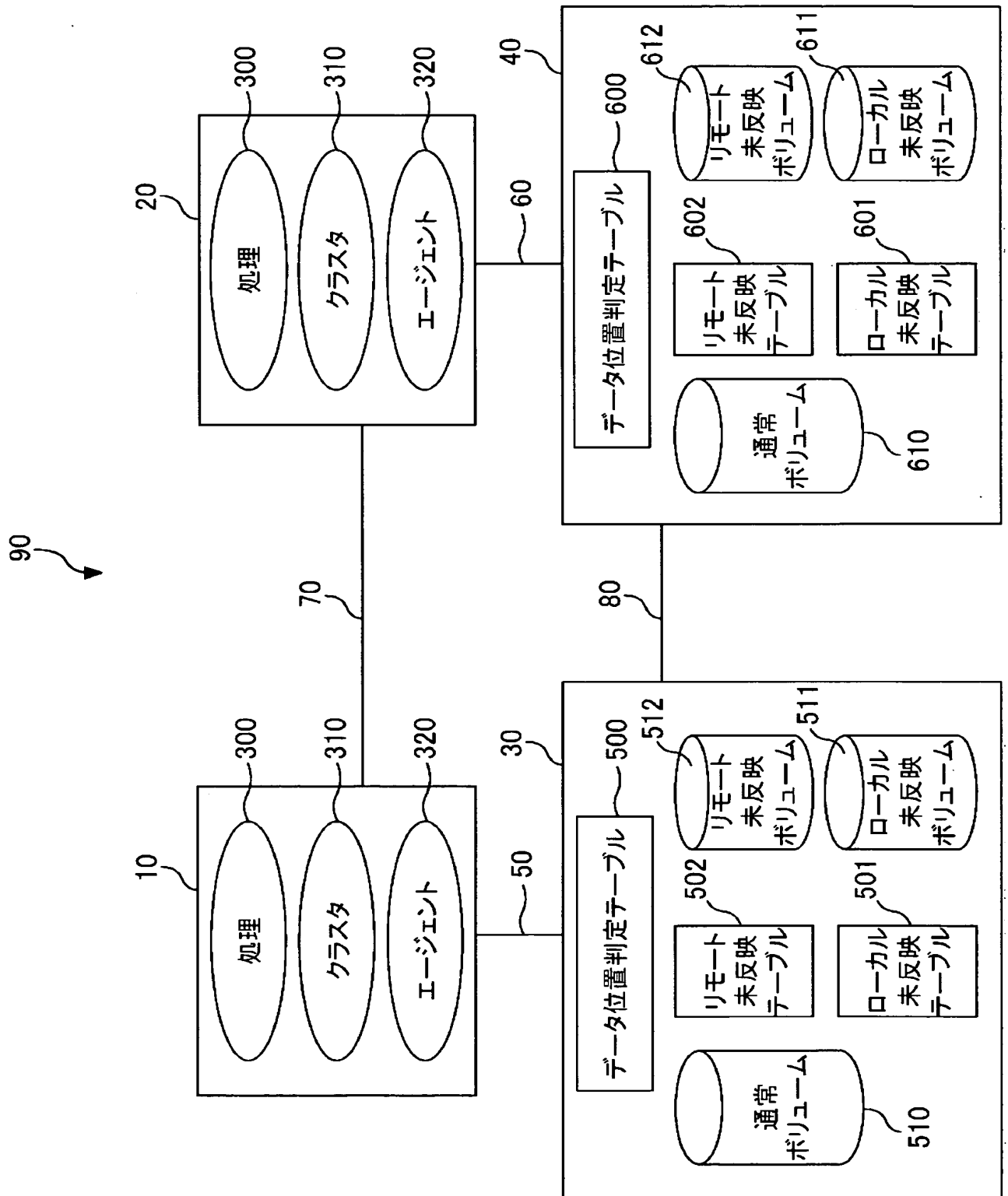
【図 9】



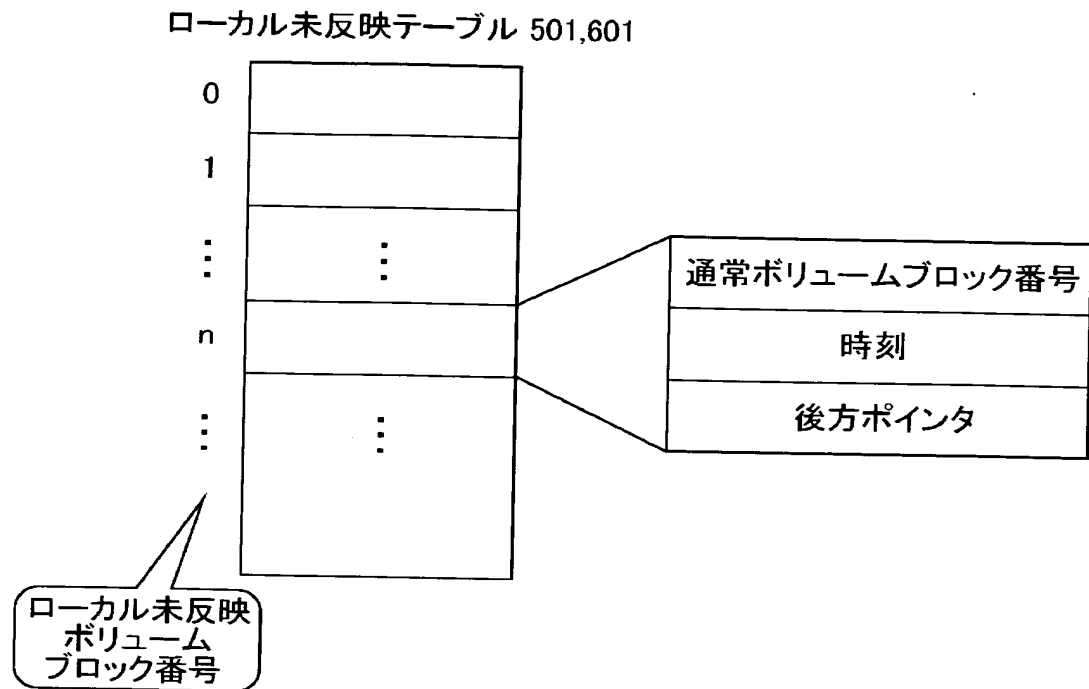
【図 10】



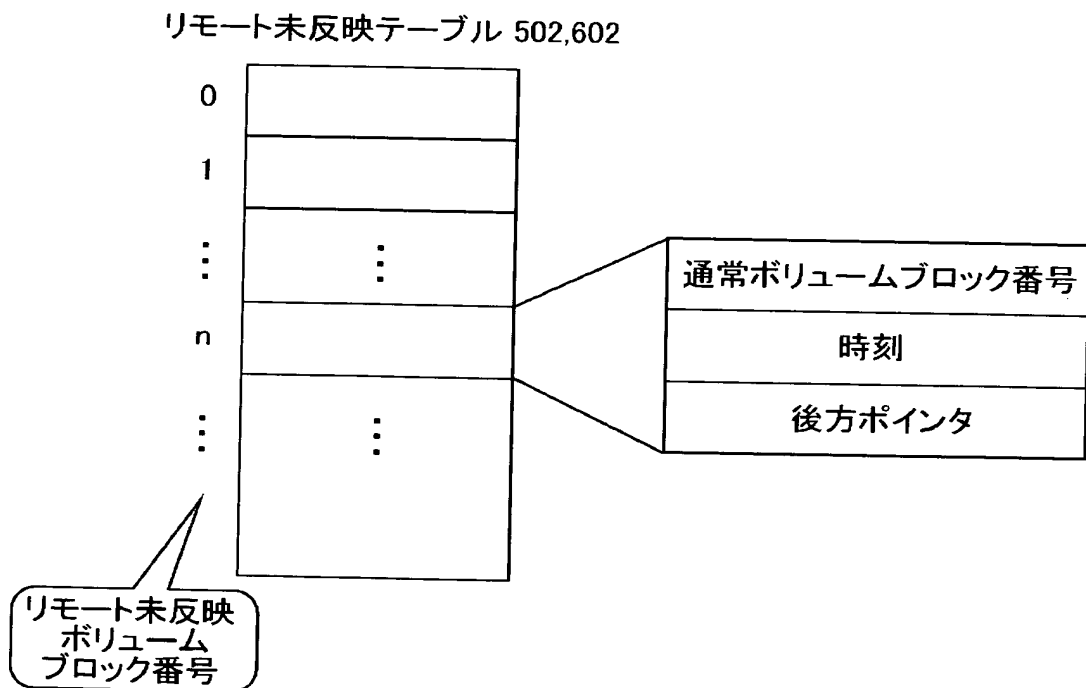
【図 11】



【図 12】



【図 13】



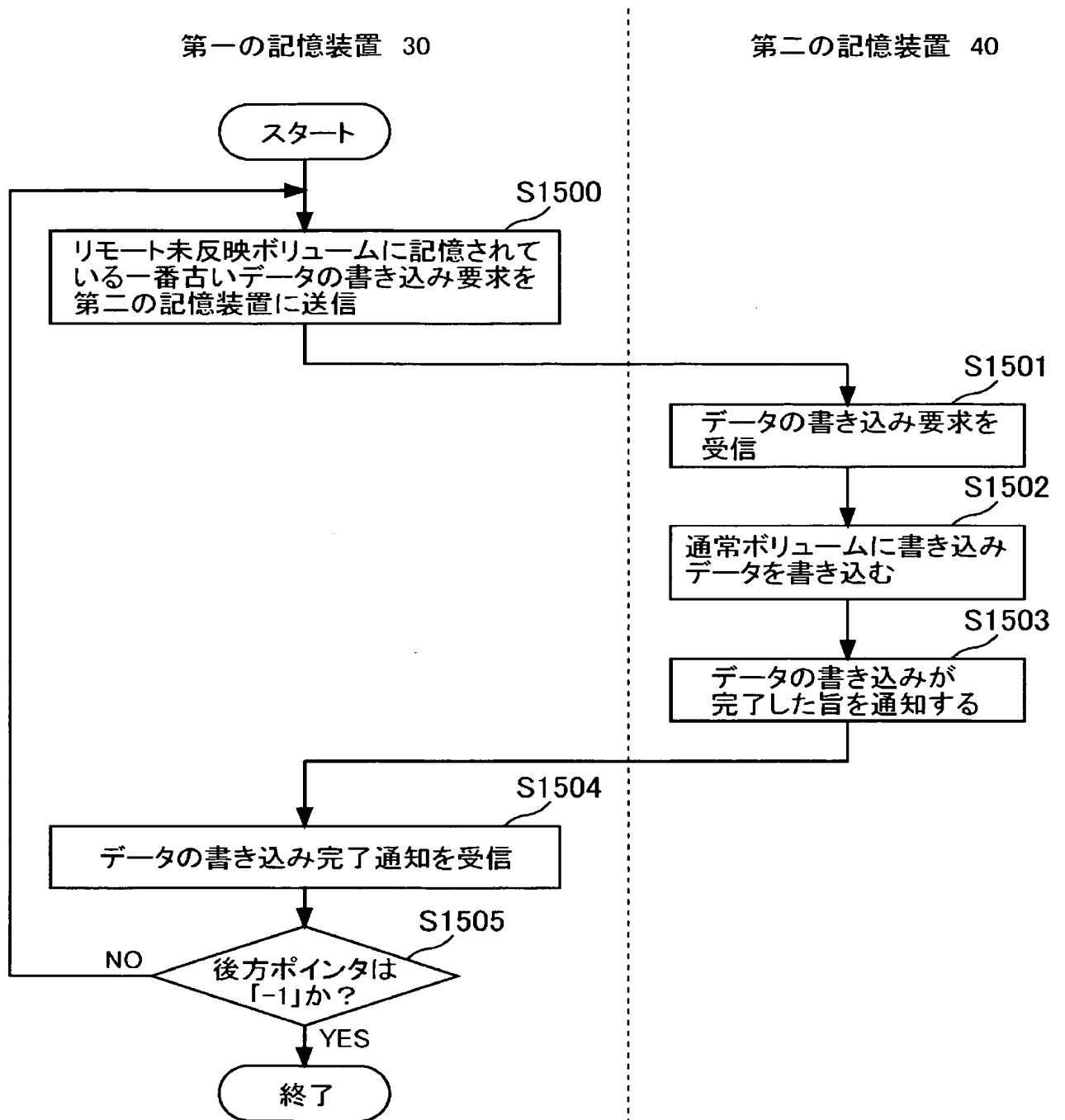
【図 14】

データ位置判定テーブル 500,600

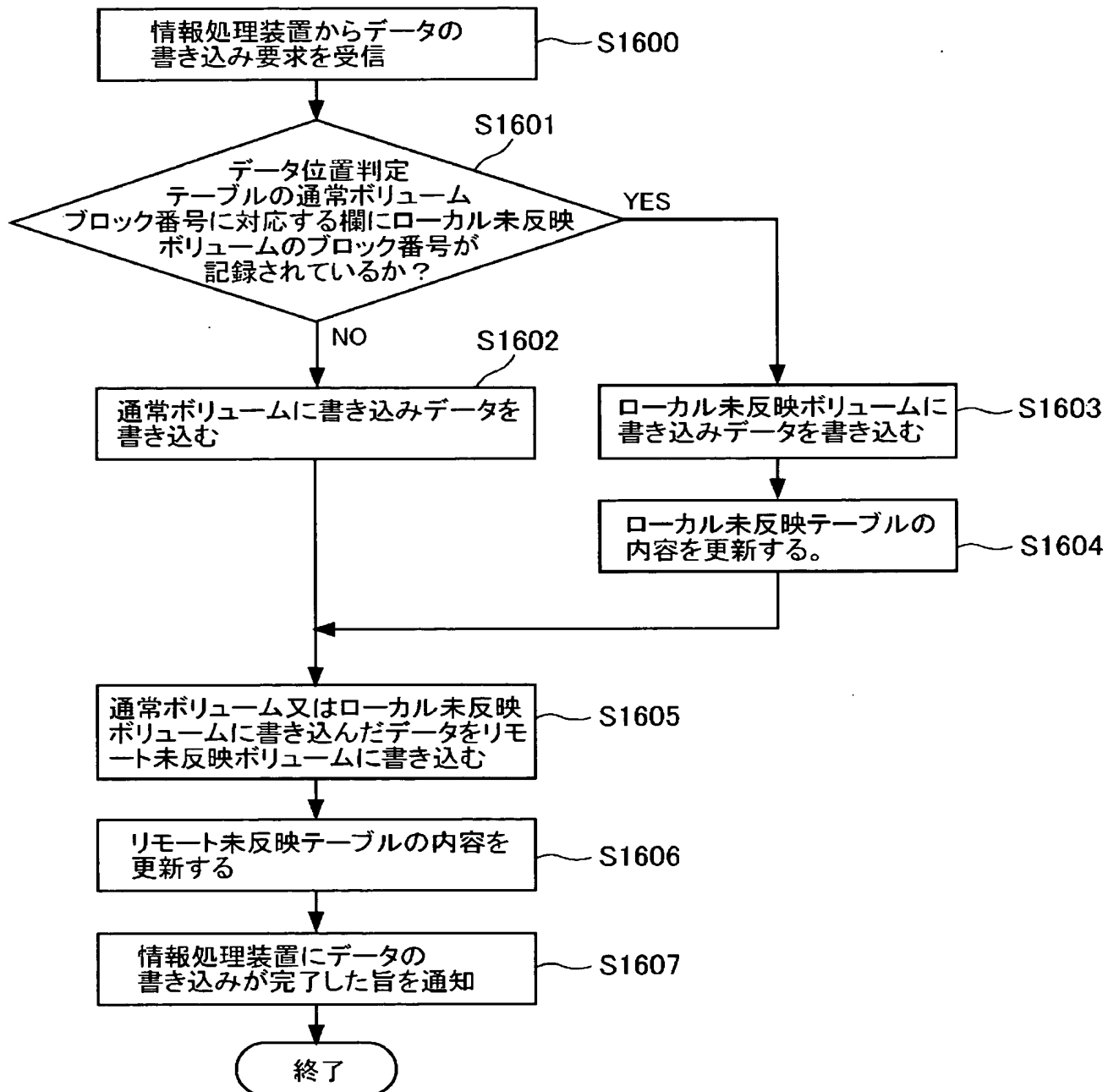
0	-1
1	-2
⋮	⋮
n	-1
⋮	⋮

通常ボリューム
ブロック番号

【図 15】



【図 16】



【書類名】 要約書**【要約】**

【解決手段】 リモートコピーの実行中に、前記第二の情報処理装置からフェールオーバーする旨の通知を受信すると、前記第二の記憶装置は、前記第一の記憶領域に書き込んだデータの複製が前記第二の記憶装置に未だ送信されておらず、前記データの複製が前記第二の記憶領域に書き込まれていないことを示す第一の情報を前記第一の記憶装置に要求し、当該要求に応じた前記第一の記憶装置から前記第一の情報を受信すると、前記第二の情報処理装置にデータ入出力要求を受け付けることができる旨を通知し、フェールオーバーした前記第二の情報処理装置から送信されてくるデータの読み出し要求を受信すると、前記第一の情報を参照して前記データの読み出し要求の対象となるデータが前記第一の記憶領域に記憶されているかを判断することとする。

【選択図】 図 8

特願 2 0 0 3 - 4 0 2 9 9 4

ページ： 1/E

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1. 変更年月日
[変更理由]

住 所
氏 名

1 9 9 0 年 8 月 3 1 日

新規登録

東京都千代田区神田駿河台 4 丁目 6 番地
株式会社日立製作所